现 代 遗 传 学 丛 书

分子遗传学

Molecular Genetics

张玉静 主编

图书馆

科学出版社

现代遗传学丛书

分子遗传学

张玉静　主编

科学出版社

2000

内 容 简 介

本书是“现代遗传学丛书”之一，图文并茂，内容翔实，从遗传的细胞学基础、遗传的物质基础、基因的复制、基因的表达、基因表达的调控、基因重组、转座 、DNA 损伤与修复、基因突变、遗传与进化等现代分子遗传学的理论进行系统深入的论述，对分子遗传学理论的重要进展有较充分的反映。

本书可供综合性大学和医学农学类大专院校分子生物学、生物化学、遗传学、生物工程学方面的教师、研究生、大学生，以及有关科研人员参考。

图书在版编目（CIP）数据

分子遗传学/张玉静主编．－北京：科学出版社，2000.4

ISBN 7-03-008038-6

Ⅰ.分… Ⅱ.张… Ⅲ.分子遗传学 Ⅳ.Q75

中国版本图书馆 CIP 数据核字（1999）第 66027 号

科学出版社 出版

北京东黄城根北街 16 号

邮政编码：100717

科地亚印刷厂印刷

新华书店北京发行所发行 各地新华书店经售

2000 年 4 月第 一 版 开本：787×1092 1/16

2000 年 4 月第一次印刷 印张：31 1/4

印数：1—4 330 字数：720 000

定价：64.00 元

（如有印装质量问题，我社负责调换（新欣））

序

分子遗传学是遗传学的一个分支，着重在分子水平上研究基因的结构和功能以阐明生物遗传和变异的规律。

基因始终是遗传学研究的对象，1909 年丹麦遗传学家约翰逊首次提出基因这一术语时，当时基因还只是代表某个遗传性状的一个抽象符号。随着研究工作的深入，不断地赋予基因以新的科学内涵。现已确定基因是一种化学分子，是遗传信息的物质载体，蕴含着支配生命活动的指令和构建生物体的蓝图，同时也是可以人工操作的改变生物体属性的一种元件。总之，所有生物体的所有生命活动无不直接或间接地在基因的控制之下。因此，凡是以生物体某一特定生命现象为研究对象的学科，如研究细胞内生命活动的细胞生物学，研究生物体生长发育的发育生物学，研究生物系统发生和进化的进化生物学，研究生物体内生理、代谢的生理学和生物化学，以及研究生物体不同层次上的神经活动的神经生物学等，在探究这些生命活动的底蕴和机制时，都会涉及基因。基因融入了生命科学的各个学科，分子遗传学已成为生命科学领域中十分重要的一门基础学科。

在分子遗传学基础上发展起来的基因工程技术，在推动整个生命科学的基础研究，促进医学、农学、环境保护工程和制药工业的发展中发挥越来越重要的作用。因此，分子遗传学知识对学生、教师、科学工作者，以及工程技术人员都是必要的。

目前，全国许多高校都开设了分子遗传学课及基因工程课。从总体上看，基因工程技术有关的各种专著较多，但并不全都是有特色，且能全面反映当前国际上分子遗传学的新进展的。本书的出版为高等学校的教学和相关的科学研究者提供了新的参考资料。本作者结合其自身的教学经验和科研实践，经过多年的努力，查阅了大量国内外新发表的文献资料进行编撰。本书内容翔实，结构合理。第一章为遗传的细胞学基础，它为更好地理解分子遗传学的基本内容打下坚实基础。第二、三两章为遗传的物质基础，详述遗传物质的本质、结构和功能，基因与基因组概念和鉴定方法。第四至八章按遗传信息传递过程进行编排，详细介绍基因复制、表达及表达调控的规律。第九至十二章阐述遗传重组、转座、DNA 的重排与扩增及免疫多样性产生的机制，揭示遗传信息的流动性和可变性。第十三章阐述 DNA 的损伤、修复和基因突变的机制。第十四章阐述遗传与进化的规律。本书系统而全面地在读者面前展示遗传物质的本质、生命的起源、发生、发展和进化的全过程。

全书内容丰富、文笔流畅、布局合理，而且逻辑性强、图文并茂，全面阐述当前分子遗传学的基本理论，并反映出该学科的前沿和发展趋势，确是一本很好的教材。我很高兴地向读者推荐。

赵寿元
复旦大学遗传学研究所
1999.11.18

主　编　　张玉静

副主编　　欧阳红生

编著者(以姓氏笔画为序)

阮承迈

吴东林

张永亮

张玉静

李慎涛

陈守义

欧阳红生

赵志辉

谭　芳

目　录

序

第一章　遗传的细胞学基础 …… (1)

第一节　原核生物和真核生物 …… (1)

一、原核生物和真核生物的概念 …… (1)

二、原核细胞与真核细胞的区别 …… (1)

三、原核生物和真核生物的起源 …… (2)

四、生物进化的二界论与三界论 …… (2)

第二节　细胞的结构和功能 …… (4)

一、细胞膜 …… (4)

二、细胞质 …… (5)

三、细胞核 …… (8)

第三节　染色体 …… (9)

一、染色体的形态特征 …… (9)

二、染色体的超微结构 …… (13)

三、染色体的大小和数目 …… (21)

四、染色体分带 …… (22)

五、核型与核型分析 …… (25)

第四节　细胞分裂 …… (26)

一、无丝分裂 …… (27)

二、细胞周期 …… (27)

三、有丝分裂 …… (27)

四、减数分裂 …… (31)

第五节　染色体畸变 …… (38)

一、染色体结构变异 …… (38)

二、染色体的数目变异 …… (42)

第二章　遗传的物质基础 …… (47)

第一节　遗传物质的本质 …… (47)

一、DNA 是遗传物质 …… (47)

二、RNA 也可以作为遗传物质 …… (49)

三、是否存在核酸之外的其他遗传物质 …… (49)

第二节　核酸的化学组成 …… (52)

第三节　DNA 的二级结构 …… (54)

一、DNA 双螺旋模型的提出 …… (54)

二、维持 DNA 双螺旋的力 …… (55)

三、双螺旋结构的基本形式 …… (56)
四、变性和复性 …… (57)
第四节 二级结构的其他形式 …… (59)
一、单链核酸形成的二级结构 …… (59)
二、反向重复与二级结构 …… (61)
三、DNA 四链结构 …… (62)
四、三螺旋 DNA …… (64)
第五节 DNA 的超螺旋结构 …… (67)
一、超螺旋结构 …… (67)
二、超螺旋影响双螺旋的结构 …… (68)
三、拓扑异构酶 …… (69)
第三章 基因与基因组 …… (72)
第一节 基因概念的历史演变 …… (72)
第二节 DNA 与基因 …… (74)
第三节 真核生物的割裂基因 …… (76)
一、割裂基因的发现 …… (76)
二、割裂基因的分布 …… (77)
三、割裂基因的性质 …… (77)
第四节 基因大小 …… (78)
第五节 重叠基因 …… (79)
一、原核生物的重叠基因 …… (79)
二、真核生物的重叠基因 …… (80)
第六节 真核生物的基因组 …… (81)
一、真核生物的基因组 …… (81)
二、基因组大小与 C 值矛盾 …… (82)
三、基因组的基因数目 …… (83)
四、必需基因和基因总数 …… (83)
第七节 真核生物 DNA 序列组织 …… (84)
一、真核生物 DNA 的复性动力学 …… (84)
二、真核生物的单一序列 …… (87)
三、真核生物的重复序列 …… (87)
第八节 细胞器基因组 …… (90)
第九节 基因鉴定 …… (92)
一、与 RNA 或 cDNA 杂交 …… (92)
二、Zoo-blot 杂交 …… (92)
三、外显子捕捉 …… (93)
四、CG 岛鉴定 …… (93)
五、DNA 序列的计算机分析鉴定 …… (93)
六、扣除杂交 …… (93)

第十节　人类基因组计划 …… (94)
一、人类 DNA 序列测定 …… (94)
二、测序技术的发展 …… (95)
三、人类基因组序列变异的研究 …… (95)
四、功能遗传学的研究方法 …… (96)
五、比较遗传学研究 …… (96)
六、伦理、法律和社会问题的研究 …… (96)
七、生物信息学和计算机生物学 …… (97)
八、遗传学研究专门人才的教育训练 …… (97)
第四章　复制 …… (98)
第一节　复制概况 …… (98)
一、半保留复制 …… (98)
二、复制起点、方向和方式 …… (100)
三、DNA 聚合酶与 DNA 聚合反应 …… (102)
四、复制的基本过程 …… (103)
五、DNA 复制的半不连续性 …… (103)
第二节　复制体系 …… (105)
一、复制体系的鉴定 …… (105)
二、DNA 聚合酶 …… (105)
三、DNA 连接酶 …… (113)
四、螺旋酶 …… (113)
五、DNA 拓扑异构酶 …… (115)
六、单链结合蛋白 …… (116)
第三节　复制的起始 …… (117)
一、预引发体的形成 …… (117)
二、复制复合物的改型 …… (120)
三、引发体的形成 …… (120)
四、复制体的组装 …… (121)
第四节　延伸过程 …… (124)
一、后随链合成的起始 …… (124)
二、前导链和后随链的协同合成 …… (125)
三、全酶的循环 …… (126)
四、冈崎片段的连接 …… (128)
第五节　复制终止 …… (129)
一、大肠杆菌 *ter*-Tus 复合物 …… (130)
二、抑制机制 …… (130)
三、复制终止体系的生物学功能 …… (131)
四、线性 DNA 的末端复制 …… (131)
第六节　其他原核生物的复制体系 …… (134)

一、噬菌体λ的复制 …… (134)
二、噬菌体 T7 复制起始 …… (135)
三、噬菌体 T4 的复制体系 …… (136)
第七节 其他形式的复制 …… (137)
一、线粒体的 D 环复制 …… (137)
二、滚环复制 …… (141)
第八节 真核生物的复制过程 …… (142)
一、概况 …… (142)
二、DNA 聚合酶 …… (142)
三、SV40 DNA 复制 …… (144)
四、酵母复制 …… (147)
五、真核生物复制时的核小体结构 …… (150)
六、真核生物染色体末端的复制与端粒酶 …… (151)
第五章 转 录 …… (155)
第一节 转录酶和转录因子 …… (155)
一、原核生物的 RNA 聚合酶 …… (155)
二、真核生物的 RNA 聚合酶 …… (159)
第二节 启动子 …… (160)
一、原核生物的启动子 …… (160)
二、真核生物的启动子 …… (164)
第三节 终止子 …… (168)
一、原核生物的终止子 …… (168)
二、真核生物的终止子 …… (169)
第四节 转录的机制 …… (170)
一、转录的起始 …… (170)
二、转录的延伸 …… (174)
三、转录的终止 …… (175)
第五节 转录产物的后加工 …… (177)
一、原核生物转录产物的后加工 …… (177)
二、真核生物的转录后加工 …… (179)
第六节 RNA 的剪接 …… (182)
一、核基因 mRNA 的剪接 …… (184)
二、第一类内含子的剪接 …… (191)
三、第二类内含子的剪接 …… (194)
四、酵母 tRNA 的剪接 …… (195)
第六章 蛋白质生物合成 …… (199)
第一节 遗传密码 …… (199)
一、遗传密码 …… (199)
二、密码子与反密码子的相互作用 …… (200)

第二节　tRNA的功能 …… (202)
一、氨酰tRNA合成酶 …… (202)
二、抑制tRNA …… (204)
三、tRNA可影响可读框 …… (206)
第三节　核糖体的结构 …… (207)
一、核糖体的构成 …… (207)
二、核糖体的装配 …… (209)
三、核糖体的活性位点 …… (210)
四、核糖体在蛋白质合成中的作用 …… (211)
第四节　原核生物的翻译过程 …… (213)
一、起始 …… (213)
二、肽链的延伸 …… (216)
三、合成的终止 …… (218)
第五节　真核生物的蛋白质生物合成过程 …… (218)
一、合成起始 …… (218)
二、肽链的延伸 …… (222)
三、合成终止 …… (223)
第六节　蛋白质生物合成初始产物的后加工 …… (225)
一、肽链中氨基酸残基的化学修饰 …… (225)
二、肽链N端甲硫氨酸或甲酰甲硫氨酸的除去 …… (225)
三、信号肽的切除 …… (225)
四、肽链的折叠 …… (226)
五、切除前体中功能不必需肽段 …… (226)
六、二硫键的形成 …… (227)
七、多肽链N端和C端的修饰 …… (227)
第七章　原核生物基因表达调控 …… (229)
第一节　概述 …… (229)
第二节　操纵子 …… (231)
一、乳糖操纵子的结构 …… (231)
二、小分子诱导物对阻抑蛋白活性的影响 …… (232)
三、操纵子突变 …… (234)
四、阻抑蛋白作用机制 …… (237)
五、蛋白质与DNA的特异结合作用 …… (241)
六、色氨酸操纵子控制模式——多基因座同时阻抑 …… (243)
七、正调控和负调控 …… (244)
八、分解代谢产物阻抑作用对启动子的正调控 …… (246)
第三节　转录后加工的调控 …… (249)
一、经mRNA切割进行的调控 …… (249)
二、通过切割释放原核和真核rRNA …… (250)

第四节　翻译水平的调控 …… (252)
一、翻译过程中的自体调控 …… (252)
二、应急调控 …… (256)
三、弱化作用 …… (258)
四、小 RNA 分子调节翻译 …… (261)
第八章　真核生物基因表达调控 …… (264)
第一节　概述 …… (264)
第二节　真核生物基因转录调控 …… (264)
一、基因转录的顺式调控元件 …… (264)
二、通用调控中的调控效应元件 …… (267)
三、DNA 结合结构域 …… (269)
四、锌指基序——一种 DNA 结合结构域 …… (271)
五、类固醇受体的几个独立结构域 …… (272)
六、结合相关靶 DNA 的同源域 …… (275)
七、螺旋-环-螺旋结构 …… (276)
八、亮氨酸拉链结构 …… (278)
九、基因激活占先模型的动力学 …… (280)
十、基因表达与去甲基化的关系 …… (281)
十一、甲基化确定印记 …… (281)
第三节　mRNA 前体的可变剪接 …… (283)
一、可变剪接的方式 …… (283)
二、可变剪接的调控机制 …… (283)
三、反式剪接 …… (284)
第四节　RNA 编辑 …… (285)
一、核苷酸的替换 …… (285)
二、可读框的改变 …… (286)
三、向导 RNA …… (286)
第五节　mRNA 稳定性的调控 …… (286)
第六节　翻译的调控 …… (287)
一、翻译因子磷酸化调控 …… (287)
二、酵母 GCN4 mRNA 翻译的调控 …… (288)
三、转铁蛋白 mRNA 翻译的调控 …… (289)
第九章　遗传重组 …… (290)
第一节　概述 …… (290)
第二节　同源性重组 …… (292)
一、同源性重组是重要而复杂的过程 …… (292)
二、进行同源性重组的基本条件 …… (292)
三、同源性重组的分子模型 …… (293)
四、联会复合体 …… (298)

五、基因转换 …… (300)
第三节 大肠杆菌重组的分子基础 …… (302)
一、*Chi* 位点和 RecBCD 核酸酶 …… (303)
二、RecA 蛋白和 Holliday 连接体的形成 …… (304)
三、Ruv 和 RecG 蛋白和 Holliday 连接体的拆分 …… (306)
第四节 DNA 的拓扑学 …… (308)
第五节 位点特异性重组 …… (310)
一、λ 噬菌体 DNA 的整合与切除 …… (311)
二、λ 噬菌体 DNA 的整合机制 …… (312)
第六节 异常重组 …… (314)
一、末端连接 …… (314)
二、链滑动 …… (315)
三、二级结构促进异常重组 …… (318)
第十章 转 座 …… (322)
第一节 概述 …… (322)
第二节 转座子的分类 …… (323)
一、插入序列 …… (324)
二、复合型转座子 …… (325)
第三节 转座发生的机制 …… (327)
一、三种不同类型的转座 …… (328)
二、转座的共同中间体 …… (330)
二、复制型转座的基本过程 …… (332)
四、非复制型转座 …… (333)
五、Tn*A* 的转座 …… (334)
六、Tn*10* 的转座 …… (335)
第四节 玉米的转座元件 …… (337)
一、玉米的控制元件 …… (337)
二、玉米的转座子家族 …… (339)
三、*Spm* 元件影响基因表达 …… (341)
第五节 果蝇的转座元件 …… (342)
第六节 反转录病毒和反转座子 …… (345)
一、概述 …… (345)
二、反转录病毒和反转座子 …… (346)
三、反转座子的分类 …… (353)
第十一章 DNA 的重排和扩增 …… (362)
第一节 概述 …… (362)
第二节 酵母接合型的转变 …… (363)
一、酵母的接合型 …… (363)
二、酵母接合型的转变 …… (366)

三、接合型相关基因的抑制与活化 …… (369)
四、接合型转变的机制 …… (371)
五、*HO* 基因表达的调控 …… (373)
第三节 锥虫抗原的变异 …… (374)
一、锥虫的表面抗原 …… (374)
二、锥虫表面抗原的变异 …… (375)
第四节 DNA 的扩增 …… (378)
一、基因组某些序列的扩增 …… (378)
二、引入外源性 DNA 到活细胞和动物体内 …… (380)
第十二章 免疫多样性产生的机制 …… (386)
第一节 概述 …… (386)
一、免疫球蛋白形成的理论 …… (386)
二、免疫球蛋白的变异性 …… (387)
第二节 免疫球蛋白基因的重组和体细胞突变 …… (390)
一、轻链基因 *V* 和 *J* 片段的重组 …… (390)
二、重链基因 *V*、*D* 和 *J* 片段的重组 …… (392)
三、*V*、*D* 和 *J* 基因两侧的序列指导基因片段的连接 …… (392)
四、*V*、*D* 和 *J* 基因片段的连接部位可以有轻度变化 …… (394)
五、体细胞突变 …… (396)
第三节 免疫球蛋白重链恒定区基因 …… (397)
一、免疫球蛋白类型的转换 …… (398)
二、膜型和分泌型免疫球蛋白的产生 …… (399)
三、免疫球蛋白产生的调节 …… (399)
第四节 T 细胞抗原受体基因 …… (402)
一、T 细胞抗原受体基因 …… (402)
二、TCR 多样性的程度 …… (404)
第十三章 DNA 损伤、修复和基因突变 …… (405)
第一节 DNA 损伤 …… (405)
一、自发性损伤 …… (405)
二、物理因素引起的损伤 …… (406)
三、化学因素引起的 DNA 损伤 …… (407)
第二节 DNA 修复 …… (408)
一、切除修复 …… (409)
二、直接（回复）修复 …… (412)
三、错配修复 …… (413)
四、交联修复 …… (414)
五、双链断裂修复 …… (414)
六、转录和修复 …… (414)
七、修饰和限制系统 …… (415)

八、大肠杆菌的挽回系统 …… (421)
九、SOS 系统 …… (423)
十、与修复有关的人类疾病 …… (425)
第三节 基因突变 …… (426)
一、突变的命名和表示方法 …… (426)
二、突变分类 …… (427)
三、基因突变的后果 …… (430)
四、基因突变方法在研究中的应用 …… (434)
第十四章 遗传与进化 …… (438)
第一节 遗传变异与进化 …… (438)
一、遗传变异的起源 …… (439)
二、基因突变与进化 …… (439)
三、染色体畸变与进化 …… (440)
第二节 DNA 的进化变异 …… (442)
一、DNA 的含量与进化机制 …… (442)
二、DNA 重复序列的进化 …… (444)
三、卫星 DNA 的等级结构及其起源和进化 …… (445)
四、“无功能”DNA …… (448)
五、DNA 序列插入和基因调控的进化变异 …… (448)
第三节 基因的分子进化 …… (450)
一、基因重复对基因进化的影响 …… (450)
二、外显子重排对基因进化的影响 …… (452)
三、不均等交换对基因进化的影响 …… (452)
第四节 基因组的起源和进化 …… (454)
一、核（类核）基因组的结构与进化 …… (454)
二、线粒体基因组的结构与进化 …… (456)
三、展望 …… (458)
第五节 分子进化 …… (459)
一、血红蛋白的分子进化 …… (459)
二、细胞色素 c 的进化 …… (461)
三、催乳激素的进化 …… (462)
四、核糖体 rRNA 的分子进化 …… (462)
第六节 群体遗传与进化 …… (463)
一、进化理论 …… (464)
二、进化的机制 …… (467)
三、物种的形成 …… (468)
第七节 展望 …… (469)
一、人类进化 …… (470)
二、生物进化工程 …… (471)
三、进化研究的新理论对生物进化工程的影响 …… (474)
汉英对照索引 …… (477)

第一章　遗传的细胞学基础

第一节　原核生物和真核生物[1~24]

一、原核生物和真核生物的概念

根据细胞的结构和遗传物质在细胞内的分布，可将生命有机体划分为原核生物和真核生物两大类。原核生物的细胞没有真正的细胞核，遗传物质存在于整个细胞之中，有时虽然有相对集中的核区，但并无核膜围绕。真核生物的遗传物质集中在有核膜包围的细胞核中，并与特定的蛋白质相结合，经过一定的等级结构形成染色体。原核生物的遗传物质只以裸露的核酸分子方式存在，虽与少量的蛋白质结合，但没有真核生物染色体那样的等级结构。习惯上，原核生物的核酸分子也称为染色体。

某些有机体，如噬菌体和病毒，既不是原核生物，也不是真核生物。它们是一种超分子的亚细胞生命形式，它们的繁殖必须在宿主体内进行，因而其遗传机制与宿主密切相关。

二、原核细胞与真核细胞的区别

原核生物与真核生物在细胞结构、功能及遗传、变异、分化等方面具有明显的差异。原核生物是单细胞的，但单细胞生物并非都是原核生物。例如单细胞酵母菌的蛋白质合成与单细胞大肠杆菌的蛋白质合成方式迥然不同，而与极其复杂的高等动物相似，因面酵母菌是真核生物。单细胞生物（原核或真核）具有自身存活及繁殖所需的全部功能，其功能依靠单个细胞来完成，面多细胞有机体（真核生物）则需通过不同组织和器官、不同细胞的协作来完成各种功能。

原核细胞与真核细胞的差异主要有以下几点：

（1）细胞结构　原核生物的细胞结构具有肽聚糖细胞壁，真核生物不具有这样的细胞壁；在胞质内，原核细胞与真核细胞内均含有核糖体，但其核糖体的分子结构有很大差异，其余细胞器如线粒体、叶绿体（植物）、高尔基体、内质网等存在于真核细胞中而原核细胞中没有；在细胞核结构方面，真核生物具有完整的细胞核结构，原核生物却不具有核膜和核仁。

（2）遗传物质的组织　原核生物只有一条染色体，而真核生物一般有一条以上的染色体；从遗传组织方面，真核生物的遗传物质与特定组蛋白结合，并经过严格的等级结构面形成染色体，原核生物的遗传物质是裸露的核酸分子或与少量特殊蛋白质结合，且没有核小体结构；在遗传物质的交换与重组方面，真核生物通过雌雄配子融合形成合子并通过细胞分裂来完成遗传物质的交换和重组，而原核生物只是通过质粒介导来实现单向的遗传物质的交换和重组。

(3) 细胞的其他功能　由于原核细胞无溶酶体，因此不能通过吞噬和胞饮作用来进行异物的消化作用；原核细胞的电子传递部位在细胞膜，而真核细胞的电子传递部位在线粒体膜；另外真核细胞内具有胞质流动而原核细胞则没有。

上述差异只是原核细胞与真核细胞在细胞水平的差别，在分子水平，原核细胞与真核细胞还具有明显的不同，如基因的序列组织、遗传物质的复制以及基因的结构、表达方式、产物修饰、调控等方面均各有特点，具体内容见有关章节。

三、原核生物和真核生物的起源

虽然原核细胞和真核细胞在细胞水平和分子水平上具有明显的差异，但所有有机体，不管是真核生物还是原核生物，以及单细胞生物或多细胞生物，都是从原始细胞(primitive cell，progenote)进化而来。虽然现在人类对生命进化的早期阶段还缺乏了解，但是作为生命起源的原始细胞，至少有两点我们可以肯定：

1) 必须具有某种形式的自我复制材料(也许是核酸)，来保证细胞完成自我复制的整个过程。这种材料后来进化成具有遗传信息的遗传物质。

2) 必须具有包被材料，使含有遗传信息的物质及原始细胞的其他成分保留在一定的范围内，以防止它们扩散到原始液质(primeval soup)中，这种包被后来进化成为细胞膜以及细胞内的其他膜系统(核膜、内质网、高尔基复合体等)。

另外从遗传物质的本质(核酸)及遗传信息表达的基本路线(复制、转录、翻译)来看，原核生物与真核生物是完全相同的。因此当描述两种生物遗传信息的传递如何保证两者细胞的遗传一致性时，我们可以认为，尽管真核生物遗传信息阅读具有许多复杂步骤，但两者在进化过程中遵循了相同的遗传法则。

四、生物进化的二界论与三界论

原核生物和真核生物的二界论分类方法，是比较经典的分类方法。两界论认为所有生命均可划分为原核生物和真核生物，而且认为今天的原核生物更接近于祖先(原始细胞)，因而真核生物是从古代真核生物进化而来的。在这一步步的进化过程中，关键的步骤是核膜的出现，细胞壁的消失，细胞体积的增大以及细胞器的获得等。

然而自 20 世纪 70 年代后期开始，两界论受到了极大的挑战，它可能会逐渐被三界论所取代。三界论并不否认原核生物和真核生物的划分，只是由于某些事实并结合分子生物学上的新发现，对生命进化提出了新的见解。

在原核生物各种属中，发现了许多巨大的差异。首先在生活环境上有巨大的差异。如一种嗜热嗜酸菌(*Sulifolobus*)能够生活在 90℃ 的水中和强酸(pH 2)的环境中，有一种嗜盐菌(*Halobacterium*)需生活在饱和盐水中。其次在代谢方面也有很大的差异。如兼性厌氧菌有无氧气均能生存，而绝对厌氧菌在有空气时是不能生存的。在氧化磷酸化过程中，有些细胞通过无氧酵解和有氧的氧化磷酸化来供能，食物中的氢原子被用来还原氧而成为水，而厌氧的甲烷菌则用氢来还原二氧化碳生成甲烷来产生能量，硫细菌则用氢来还原硫成为硫化氢来供能。而且，在分子水平上，各种属的原核生物也存在着很大差

异。

1969 年 C. Woese 研究了各种古细菌的 16 S rRNA 的序列，并根据 16 S rRNA 的序列来确定各种生物间的亲缘关系。由于各种生物都有 16 S rRNA，而且含量丰富，易于分析，在进化过程高度保守，与其他方法相比较（如细胞色素 c，各种生物间有的缺乏，有的差别太大无法分析），有着很多的优势。到 1977 年，由于数据的积累和系统树的建立，C. Woese 提出了三界论，即原核生物应当分为两界：古细菌（archaebacteria）和真细菌（bacteriotida），这两界彼此不同，就像它们不同于真核生物一样。三者都起源于共同的祖先。三界论开始提出时，遭到许多生物学家的怀疑与反对，然而 Woese 却不断完善他的 16 S rRNA 系统树，同时积累其他生物学证据，现在，Woese 的三界论，越来越受到生物学家的重视。

在逐渐被接受的 Woese 系统树中，Woese 提出真细菌是第一个从古细菌和真核生物总干上分化出来的分支（图 1－1）。尽管古细菌与真细菌在表型、缺核、基因组织及代谢途径方面有明显的相似性，但真核生物与古细菌在分子进化上也有很大的共同性。真核生物与古细菌复杂多样的分子加工机制中的共同因素（如转录过程和泛素指导的蛋白水解作用）也支持了上述进化关系。Grarrett 等于 1994 年报道了古细菌 rRNA 内含子（intron）具有编码持家内切酶的功能。核仁纤维蛋白是一种在真核生物中与小核 RNA 有关并指导 rRNA 加工和甲基化的蛋白质，它在古细菌中的发现，暗示了真核生物与古细菌在 RNA 加工方面也具有共同性，并与真细菌有所差异。另外，Malene 于 1997 年指出在 tRNA 内含子剪接方面，两者的内含子从组织形成及剪接机制方面也是非常相似的。

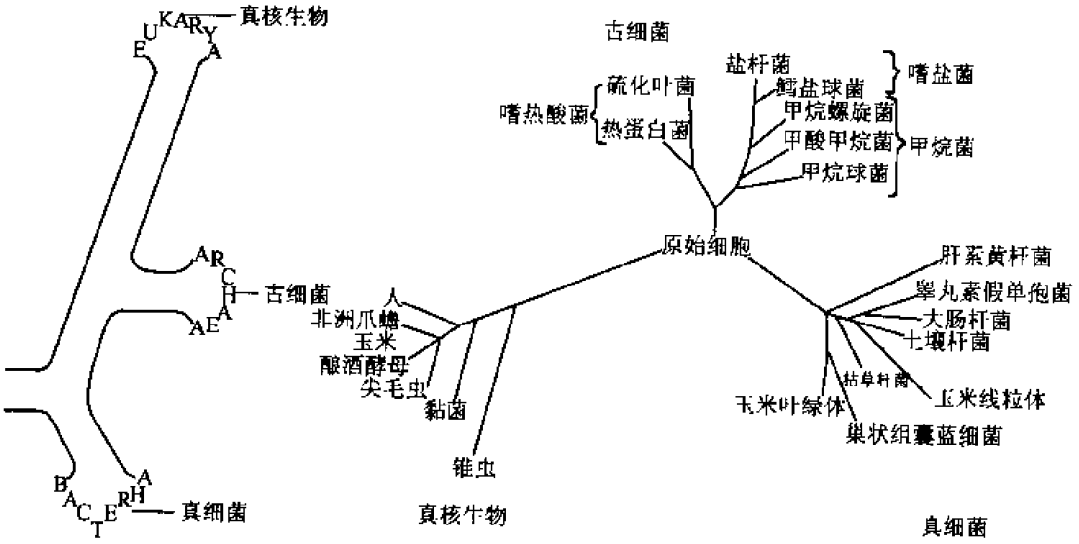


图 1－1　生物进化的途径和基于 16 S rRNA 序列分析的 Woese 系统树

随着分子生物学的发展和新证据的不断发现，生命进化的基本途径将会越来越清楚地展现在人们面前。

第二节　细胞的结构和功能[25-40]

细胞学理论创建于19世纪中期，该理论认为所有生物均由细胞所构成，对于原核生物来说，它由细胞膜所包被并与外界相隔离，遗传物质存在于细胞中形成类核(nucleoid)。真核生物的遗传物质存在于由核膜包围的细胞核中，外面由细胞质所围绕，细胞质中含有细胞器(organelle)，细胞质外是细胞膜。原核生物还具有肽聚糖的胞壁，真核生物中的植物则含有纤维素、半纤维素和果胶质构成的细胞壁。动物细胞结构一般包括细胞膜、细胞质和细胞核。

一、细　胞　膜

细胞膜(cell membrane)又称质膜(plasma membrane, plasmalemma)，是细胞的重要组成部分，它包在所有细胞细胞质的外面，为一层极薄的膜，具有独特的结构和功能。在真核细胞膜系统中，除了质膜外，膜内还有构成各种细胞器的膜，称为细胞内膜。相对于内膜，细胞膜又称为外周膜。外周膜和细胞内膜统称为生物膜(biomembrane)。细胞膜是使细胞成为生命基本单位的重要保证，其基本作用是维护细胞内环境的相对稳定，与外界环境不断地进行物质交换，能量和信息的传递，对细胞的生存、生长、分裂及分化都至关重要。

根据对各种细胞膜分析的结果表明，细胞膜主要是由脂类和蛋白质组成。有许多质膜还含有少量的糖类，形成糖脂和糖蛋白。大体上说，脂类约占总量的50%，蛋白质约占40%，糖类约占1%～10%。实际上它们的比例随膜的种类不同而有很大的差异。

膜中的脂类和蛋白质含量的变化与膜的功能有关。一般膜中含蛋白质越多，膜的功能越是复杂多样。相反膜的功能越简单，所含蛋白质的种类和数量越少。

构成膜的脂类有磷脂、胆固醇和糖脂，在膜中以磷脂最为丰富。真核生物质膜的主要磷脂有磷脂酰胆碱、磷脂酰乙醇胺、磷脂酰肌醇、磷脂酰丝氨酸和神经鞘磷脂等。磷脂分子的一个显著特点是极性很强，即有一个磷脂酰碱基的极性很强的头部和两条较长的脂肪酸链构成的非极性疏水尾部。这种一头亲水一头疏水的分子称为两亲性分子(amphipathic molecules)或称双亲媒体分子、兼性分子。磷脂分子这种独特的结构使其具有独特的物理性质，即在水环境中，磷脂分子的极性头部亲水而非极性的尾部避开水，伸向分子内部，从而形成球形的分子团(micelle)或脂质体(liposome)，或者在一定的水环境中，形成亲水头部伸向表面而疏水尾部伸向内侧的"脂质双分子层"(lipid bilayer)。大部分磷脂和糖脂分子在水环境里能自发形成双层，而且这样的脂质双层具有自我组装、自我封合的特性和流动性。此外，膜脂的亲水部分与膜蛋白的亲水氨基酸侧链结合，膜脂的疏水部分与膜蛋白疏水的氨基酸侧链进行疏水结合，从而使膜形成一种稳定的结构。膜脂对膜功能蛋白的活性主要有以下作用：①膜脂构成膜的支架，为膜蛋白(酶)维持构象和保持活性提供适宜环境。②膜上许多酶活性依赖于膜脂存在，去掉脂类，则酶蛋白失活。另外，磷脂分子可以作为在细胞膜表面的非蛋白质锚(如糖基-磷脂酰肌醇锚，GPI)，在蛋白质的跨膜运输过程及细胞膜表面的定位中起着重要作用。

对于脂溶性分子来说，如类固醇激素，可直接通过细胞膜而进入细胞内结合于细胞内的受体，对于水溶性的物质，如离子、无机分子、多肽及蛋白质等，是不通透的。这些物质的越膜转运需要细胞膜上的蛋白质来完成。膜中的蛋白质，根据其与膜脂相互作用的方式及其在膜中所在的位置不同，大体可分为外周蛋白(extrinsic 或 peripheral protein)和内在蛋白或整合蛋白(intrinsic 或 integral protein)。前者分布在膜的内外表面，为水溶性的，后者以不同程序嵌入脂双层的内部，有的还横跨全膜，它们的疏水区域与脂类分子的疏水尾部相互作用，亲水区域则暴露于一侧或两侧表面。跨膜蛋白疏水部分几乎都含有一个或一个以上的α-螺旋结构片段。细胞膜蛋白除了在水溶性物质跨膜运输方面的作用外，对于细胞的抗原性、信号识别与传递及其他功能的实现，都有着重要的作用。

二、细 胞 质

细胞膜以内、细胞核以外的细胞组成成分称为细胞质(cytoplasm)。细胞质包括基质(cytoplasmic matrix)、细胞器(organelle)和内含物等。细胞质基质是指细胞质内在细胞器和内含物以外，较为均质的半透明液态部分，它为细胞质内所进行的多种代谢反应提供内环境，并含有参加胞质内代谢反应所需的多种酶类、底物和离子，是细胞质的重要组成部分。内含物是一些胞质内除细胞器外的有形成分，有的是代谢产物，有些是储存的营养物质，如糖原、色素等。细胞器有许多种，其中具有界膜的如内质网、高尔基复合体、溶酶体、叶绿体、微体(过氧化物体)等与细胞核的核膜一起，统称为内膜系统(internal membrane system)；不具有界膜的细胞器有核蛋白体、中心体以及微管、微丝等。所有这些细胞器在基质中各有其一定的形态结构，完成一定的功能。动物细胞的细胞器主要有：

1.内质网

内质网(endoplasmic reticulum,ER)是由一层单位膜所组成的一些形状大小不同的小管、小囊或扁囊(亦称膜层 lamellae，或 cisternae，小池、潴泡)构成的。在细胞质中，一般由这些小管、小囊或扁囊连成一个连续的网状系统。内质网分为两种：含有大量核糖体颗粒的为糙面内质网(rough endoplasimic reticulum，rER)，它普遍存在子分泌蛋白质的细胞中，其数量随细胞不同而不同，越是分泌旺盛的细胞越多，新合成的分泌型蛋白质和定位于内质网、高尔基复合体、细胞膜等结构的蛋白质，必须首先从胞液(cytosol)中输入粗面内质网中，而且只有那些具有特殊疏水信号肽的蛋白质才可进入内质网，这段信号肽可以被结合于新生肽链和核糖体上的信号识别颗粒(signal-recognition particle，SRP)所识别，并进一步指导该多肽链与粗面内质网上的受体蛋白相结合。从而启动 ATP 依赖型的越膜蛋白的定位过程。因此，粗面内质网是分泌型蛋白合成、加工的场所并最终形成合适的可定位的蛋白质形式。无核糖体附着的内质网为滑面内质网(smooth endoplasmic reticulum，sER)，sER 的单位不是扁囊，面是小管小囊成分，它们连接成网，广泛存在于能合成类固醇的细胞中，它对胆固醇的代谢、类固醇激素的形成与释放有重要作用。

2.高尔基复合体

高尔基复合体(Golgi complex)，又称高尔基器(Golgi apparatus)或内网器(internal

reticular apparatus)，一般位于细胞核的附近，在动物细胞中接近中心体，其显著特征是一些扁平囊所形成的扁盘堆叠结构(Golgi stacks)，在其周围常有一些小管和大小不同的囊泡。扁囊中充满无定形或颗粒的内容物。在许多细胞中，高尔基器的扁囊堆常常明显地呈现凸面(又称顺面，*cis* face)和凹面(又称反面，*trans* face)。凸面有小囊小管连接成网状，在凹面常见大囊泡，这是分泌产物集中的地方。高尔基复合体中的糖基转移酶是其中最具特征的酶类，它能把低聚糖转移到蛋白质上形成糖蛋白。从形态和化学分析看：高尔基体的顺面膜厚度与 ER 相近，而反面膜的厚度与质膜相似，其化学成分也有相同的趋势。从膜流来看，新合成并进入内质网的分泌型蛋白在内质网腔内进行肽链折叠后，进行了初步的糖基化并形成转运小泡，从内质网来的转运小泡膜与高尔基复合体的顺面膜融合，新合成的多肽链在高尔基复合体中进行加工，完成糖基化过程并进行分拣(sorting)，然后从反面的扁囊出芽形成包被囊泡(coated vesicles)，包被囊泡的膜掺入质膜并通过形成分泌泡排出细胞。

3.线粒体

线粒体(mitochondria)较普遍地存在于各种动物细胞中。在光镜下观察，线粒体呈线状、棒状或颗粒状；在电镜下观察，线粒体为内外两层单位膜所构成的囊状结构。它的主要结构包括外膜、内膜、嵴、膜间隙和液态基质。在线粒体中，通过氧化磷酸化进行能量转换，为细胞内各项生命活动提供能量。因此，线粒体是细胞的“动力工厂”或“能量转换器”。除此之外，线粒体是动物细胞中核外唯一含有 DNA 的细胞器。

线粒体 DNA(简称 mtDNA)较小也较简单，存在于线粒体基质中，有时与线粒体内膜结合存在。在哺乳动物中线粒体基因组一般是裸露的共价闭合环状 DNA 分子，分子大小为 16.5kb 左右。线粒体 DNA 是细胞质遗传的物质基础，构成了独立的核外遗传体系(表 1-1)。该遗传体系的主要特点有以下几点：

(1)线粒体 DNA 都有编码作用　　与核基因组比较，动物几乎所有的线粒体 DNA 单核苷酸都有编码作用，或者编码蛋白质，或者编码 rRNA 和 tRNA，这些基因在线粒体 DNA 分子上排列紧凑，相邻之间没有或只有很少几个非编码碱基隔开(但酵母线粒体基因组有相当数量的非编码序列存在)。

表 1-1　人类核基因组与线粒体基因组比较

项　目	核基因组	线粒体基因组
大　小	3 000 Mb	16.6 kb
DNA 分子数	23(XX)或 24(XY)，线性	1 个环状 DNA 分子
每细胞 DNA 分子总数	单倍体 23，二倍体 46	几个 10^3
结合蛋白	几种组蛋白和非组蛋白	大部分无蛋白质
基因密度	约 1/40 kb	1/0.45 kb
重复 DNA	丰　富	几乎没有
转　录	复制单元	多基因连续转录
内含子	大部分基因均已发现	缺　乏
编码 DNA	约 3%	约 93%
重　组	减数分裂时每对同源染色体至少一次	无
遗　传	X 常染色体孟德尔遗传 Y 染色体父系遗传	绝对母系遗传

(2)线粒体具有自己的蛋白质合成体系　线粒体内有进行蛋白质生物合成所必需的各种 RNA(rRNA、tRNA 和 mRNA)。杂交试验表明,线粒体中的 RNA 是由线粒体 DNA 编码,在线粒体中合成并用于线粒体蛋白质合成。虽然未见有细胞中 RNA 的输入现象,但由细胞核 DNA 编码并在细胞质中合成的蛋白质(酶类)可输入线粒体中。如线粒体内基因转录过程中所需的 RNA 聚合酶等。

(3)线粒体基因组突变频率高　线粒体具有自身复制和分裂增殖的能力,其基因组的突变频率远远高于核基因组的突变能力。

(4)线粒体内的密码子具有更大的摇摆性　线粒体中蛋白质的合成仅需要 22 种 tRNA,除具有 6 个密码子的亮氨酸和丝氨酸有两种 tRNA 外,其余各种氨基酸密码子的识别均只有一种 tRNA 来完成,即当一个 tRNA 分子的反密码子识别 4 个一组某种特定氨基酸的密码子时,tRNA 反密码子第一位的碱基可识别密码子第三位上的任何一种碱基。

(5)线粒体中的密码子不同于通用密码子　所有生物使用着几乎完全相同的遗传密码,这也提示所有生物有着共同的祖先。然而线粒体的密码子与通用密码有明显的差异。如 UGA,不是终止密码子而是色氨酸密码子,AGA、AGG 不是精氨酸密码子而是终止密码子(哺乳动物线粒体)和丝氨酸密码子(果蝇 *Drosophila*),CUA 在酵母线粒体中不是亮氨酸密码子而是苏氨酸密码子。

(6)线粒体遗传方式属非孟德尔遗传　在有性生殖的生物中,后代的产生依赖于雌雄配子的受精作用。从细胞学角度来看,卵细胞除了细胞核之外,还有大量的细胞质以及存在其中的细胞器,而精子除了细胞核外,没有或者带有极少的细胞质,因此也就没有或极少有细胞器,所以在受精过程中,卵细胞不仅为子代提供核基因,也为后代提供了细胞质基因,使后代的某些性状受母亲的影响,从而表现出不同于核基因的非孟德尔遗传或母系遗传的遗传方式。

须指出的是,虽然线粒体基因组含有遗传信息,但其信息量非常有限(在人类的线粒体只编码 13 种多肽序列),这对于线粒体完成其能量转换和供应以及其他生命活动是远远不够的,因此大多数线粒体内蛋白质由核基因组编码,由细胞质内的核糖体合成运到线粒体内,也就是说线粒体是半自主性细胞器。

4.中心体

中心体(centrosome),常见于动物及低等植物的细胞中,位于细胞中心,常靠近细胞核。在光学显微镜下,中心体呈颗粒状。在电镜下观察,中心体由两个互相垂直的中心粒(centriole)与周围透明的、电子密度高的物质(PCM)构成。中心粒呈中空的圆筒状,筒壁由 9 组三联微管组成,一般认为中心体与细胞分裂期的纺锤体的形成及排列方向和染色体的移动有密切关系。此外,还可能参与纤毛或鞭毛的形成,而纤毛和鞭毛是某些单细胞生物和精子的运动器官。

三、细 胞 核

细胞核(nucleus)是细胞中的一个重要组成成分。细胞核的出现是进化史上的一个飞跃,由于核的出现,提高了细胞对环境的适应能力,为单细胞生物向多细胞生物进化奠定了基础。

一般真核生物细胞只有一个核,但也有两个到多个核的多核细胞,如蟾蜍的肝细胞通常有两个核,兔、鼠的肝细胞可达 10 个核。极少数的高度分化细胞没有核(如哺乳动物的成熟红细胞)。

核一般位于细胞的中央,但随着细胞的生理性状不同,其位置也会改变。如在生长的细胞中,核往往移到生长作用最旺盛的地方,这说明细胞核与细胞的代谢活动有着密切的关系。

细胞核的形态结构,在细胞生活周期的各个阶段变化很大。细胞分裂间期的细胞核称为间期核,其主要结构包括:核膜、核基质、核仁、染色质(染色体)。

1.核膜

核膜(nuclear membrane)由内外两层单位膜组成,并把胞质与核质分开。面向胞质的外膜表面附着大量的核糖体颗粒,通常可见内质网与其相通连。内膜面向核质,表面没有核糖体颗粒,但和浓集的染色质紧密接触。核膜上有许多规律排列的核孔(nuclear pore),是调节细胞核与细胞质之间物质交换的通道,并且有一定的选择性。小分子物质交换在整个核膜进行,大分子物质(如核内形成的各种 RNA)通过核孔进入细胞质。

2.核基质

核基质(nucleoplasm)是核内的无形部分,为透明胶体物质,又称核液。它的成分和细胞基质很相近,含有多种酶和无机盐。

3.核仁

在光学显微镜下观察,细胞核中有一个或几个均质的反光性很强的小体,称为核仁(nucleolus)。在电子显微镜下观察,可以看到核仁在构造上的分化。一是位于核仁周缘的颗粒区,含有直径为 15~20nm 的颗粒成分;二是位于核仁中心的纤维区,含有直径为 5~10nm 的纤维成分;三是直径为 10nm 的细丝状染色质,它位于核仁的周围并伸展到核仁内部;四是基质,为上述三种成分以外的部分。核仁最主要的功能是合成核糖体 RNA(rRNA),并与细胞质内核糖体的生物合成有关。核仁往往与染色体上的特定部位(次缢痕)相联结,因此这个染色体上的特定部位又称核仁组织区(nucleolus organizing region, NOR)。

4.染色质

染色质(chromatin)是指真核生物细胞核细胞分裂间期能被碱性染料着色的物质,是由 DNA、组蛋白、非组蛋白和少量 RNA 所组成的复合物,是细胞分裂间期遗传物质的存

在形式。在光镜下，间期核内染色质呈染色较深的细网状不规则结构，进入分裂期时则形成形态不同而又各具特征的高度浓缩的染色体。

染色质和染色体是真核生物遗传物质存在的两种不同形态，反映了它们处于细胞分裂周期的不同功能阶段，两者不存在成分上的差异。真核细胞的绝大部分 DNA 都集中在染色体上，所以染色体是遗传信息的主要载体。

总之，细胞核的主要功能是储存和复制遗传信息，在一定程度上控制着整个细胞的生命活动。在细胞核内以脱氧核糖核酸为模板，在核内有关酶类的参与下，合成核糖核酸。核糖核酸经膜孔进入细胞质中，在核糖体上按照核糖核酸的碱基序列，合成一定结构的蛋白质(酶)，从而参与细胞内的代谢、分化和繁殖等生命活动。同时细胞是有高度组织性的统一整体，细胞核与细胞质，各细胞器之间都有密切的联系。只有各部分密切配合，互相制约，才能完成整个生命活动。

第三节　染 色 体[41~78]

染色体(chromosome)原指真核生物细胞分裂中期具有一定形态特征的染色质，现在这一概念已扩大为包括原核生物及细胞器在内的基因载体的总称。大部分原核生物的染色体形态比较简单，它只是一条裸露的或与少数蛋白质结合的 DNA 或 RNA 双链或单链。真核生物染色体的形态特征，只有在核分裂过程中当染色体缩短时方可在光镜下观察到。在分裂间期的细胞中染色体为纤细的线状，称为染色线(chromonema)。染色线彼此缠绕，以致于很难区分一组染色体中的单个成员。在有丝分裂中期，这些染色线经过卷曲和高度卷曲(coiling and supercoiling)，染色体凝缩达到最大程度，成为在光镜下可识别的，具有一定形态特征的染色体。

一、染色体的形态特征

(一) 染色体的一般形态结构

通过技术处理，在光学显微镜下可观察到细胞分裂中期染色体具有以下结构特征：

1. 着丝粒和主缢痕

在光镜下观察细胞分裂中期的染色体时，都会发现一个相对不着色的缢缩部位，称为主缢痕(primary constriction)。在这个区域两侧的染色体部分称为染色体的臂(arm)。染色体臂染色较深，将主缢痕区明显地显现出来。着丝粒(centromere)是染色体上纺锤丝的附着部位。如图 1-2 所示，在电镜下观察哺乳动物染色体时发现，主缢痕两侧为一对三层结构的特化部位，即着丝粒两侧的由蛋白质构成的三层盘状或球状结构，称动粒(kinetochore)。着丝粒是主缢痕处的染色质部位，在有

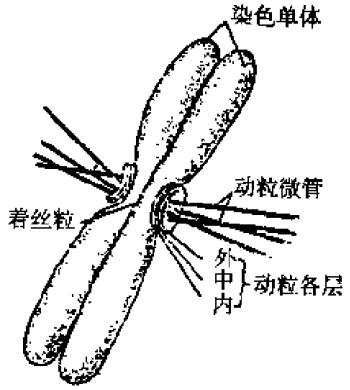


图 1-2　中期染色体上的着丝粒

丝分裂的前期和中期，着丝粒把两个姐妹染色单体连在一起，到后期，两个单体的着丝粒分开，纺锤丝把两条染色单体拉向两极，着丝粒有染色体牵丝纤维与其相连。如果染色体上没有着丝粒，中期染色体就不能在赤道板上取向排列，后期也不能向两极移动。着丝粒使染色体在细胞分裂时表现出形态上的不连续性，因此人们认为着丝粒是两个染色体臂DNA的一段特殊的高度重复的DNA序列，该序列不能与组蛋白结合，却与其他蛋白质结合形成动粒，从而与纺锤体的微管蛋白结合，来参与有丝分裂后期的染色体移动。

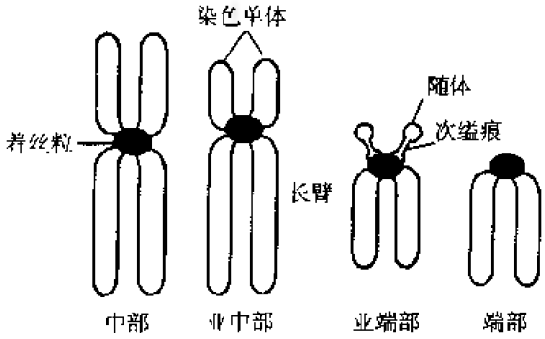


图1-3 根据着丝粒位置的染色体分类

并非有丝分裂各期或多种生物的染色体均有此种分化结构。如鸟类的大染色体有此盘状三层结构，而小染色体则没有，酵母没有明显的动粒，微管似乎直接插入染色体中。

通常每个染色体只有一个着丝粒，它在某条特定染色体上的位置也比较恒定。但不同染色体上的着丝粒的位置则不相同。所以常用它作为描述一个染色体组每条染色体的一个标记。习惯上根据着丝粒的位置将染色体分为几种类型(图1-3)：

(1)中着丝粒(metacentric)染色体　其着丝粒大约位于染色体的中部，使两个染色体臂的长度大致相等(臂比值1.0～1.7)；

(2)近中着丝粒(submetacentric)染色体　是指着丝粒的位置虽靠近中部，但与两端的距离不等，使染色体的两臂具有明显不等的长度(臂比值1.7～3.0)；

(3)近端着丝粒(subtelocentric)染色体　是指着丝粒靠近染色体的端部但距端部还有一定距离(臂比值3.0～7.0)；

(4)端着丝粒(telocentric)染色体　即着丝粒在染色体的端部，形成只有一个染色体臂的染色体(臂比值>7.0)。

由于对有无真正的端部着丝粒染色体有争论，又常使用近端着丝粒(acrocentric)染色体这个词，来称呼只有一个染色体臂的染色体。

在某些情况下，着丝粒区域的功能由一个次生的移动中心所代替，形成"新着丝粒"(neocentromere)。有的生物的染色体可以具有两个(双着丝粒染色体)或两个以上的着丝粒，结果纺锤丝可以与染色体上许多位点相连接，故称为多着丝粒(polycentromere)染色体，如马蛔虫曾发现这类例证。在一些昆虫的同翅目(Homoptera)、半翅目(Hemiptera)，沿着整个染色体长度的每一个点都具有着丝粒的活性，这种活性本质上是扩散的，故称之为弥散着丝粒(holocentromeres)。

由着丝粒的位置而形成的染色体臂，可用不同方式测量。根据1966年国际芝加哥会议确定，短臂用"p"表示，长臂用"q"表示，则臂的量比值 $A=\frac{p}{q}$，或用着丝粒指数 $\frac{p}{p+q}\times 100\%$ 计算。

2.次缢痕和随体

在染色体组中，除每一条染色体都有主缢痕外，在个别染色体上还有另外一个染色较

淡的缢缩部位，称为次缢痕(secondary constriction)。它通常出现在染色体末端的近旁，其大小和范围是恒定的，一般位于染色体的短臂上。这一区域与末期核仁形成有关，并在间期和前期与核仁联系在一起，因而被称为核仁组织区。业已证明，核仁形成区是 rRNA 的合成部位。然而在中期，核仁一般看不到，但在这一区域往往具有显著的标记，即在该部位的异固缩(heteropycnosis)程度很高，易被识别，因此是染色体组中鉴别该种染色体的良好佐证，并在分类学上有很大价值。

与次缢痕相连的圆形或椭圆形小体称为随体(satellite, trabant)。它通过随体柄连接到染色体短臂上。随体的大小差异很大，有的小到难以辨别的程度，而有的是巨型随体。此外，同源染色体的随体大小也不一定相同。一个种的核仁形成可能对另一个种为显性，所以在杂种中，将会出现一个亲本染色体组显示有随体，而另一个亲本则无随体存在。目前人们认为随体几乎全部是异染色质，所以它不具有常染色质所具有的遗传活性。然而 Phillips 等证实，在玉米中决定多次有丝分裂的 *Po* 基因位于 6 号染色体的随体上。此外还有人检测出与核仁形成区无关的次缢痕，为避免与核仁形成区相混淆，称之为第二缢痕。在很多动植物观察到的第二缢痕可能是螺旋化程度不同或核酸构成成分有差异的区域和活性薄弱的区域。随体是识别染色体的重要特征之一。

3.染色粒

在减数分裂的粗线期，染色线表现出沿着染色体全长分布的，具有恒定大小和一定位置的"念珠"状突起，称为染色粒(chromomere)。染色粒与染色粒之间着色较浅的区域为染色粒间区(interchromomere)。目前对染色粒的形成所能接受的解释是连续的 DNA 链的局部螺旋化所致。它可能相当于 DNA 的复制单位，RNA 的合成单位和加工单位。异染色质的染色粒比常染色质的染色粒染色深，似乎比常染色质的染色粒大，因此称之为大染色粒。通常靠近着丝粒的染色粒大，而向着染色体末端逐渐变小，从而构成染色粒的大小梯度。在粗线期染色体上，染色粒的数目基本是恒定的，因而可作为可靠的形态特征。

4.端粒

端粒(telomere)是染色体上增大的末端染色粒，是染色体的组成部分。像着丝粒一样，当端粒缺失时，则该染色体无正常功能。端粒能封闭正常染色体末端，具有端粒的染色体末端不能与其他断裂的染色体末端连接。假如正常染色体被 X 射线打断，形成的断片可重新愈合。然而，它们不会和染色体臂末端愈合，臂的末端也不能相互愈合。人们认为，位于染色体末端的端粒使染色体具有极性(polarity)，从而保证了染色体在细胞周期中的正常行为。在特殊情况下，端粒具有着丝粒的活性，因而被称为新着丝粒。Faria 等在很多的植物物种中研究了端粒的复合结构，以为端粒是由两个独立而又有区别的原端粒(protelomere)和真端粒(eutelomere)组成。根据观察，原端粒是具有明显界限的末端深染结构，而真端粒是连接原端粒的亚末端部分。并认为，一个复合端粒含有 8 个不同的染色粒，若其中部分断裂不会丧失该结构的遗传功能。根据电子显微镜研究，端粒是由无规则折叠的染色丝和蛋白质组成，其染色丝很少终止于染色体末端，而是反折回到染色单体中去。

5.染色单体

中期染色体由两条染色单体所组成，两者在着丝粒的部位相互结合。每一条单体是一条 DNA 双链经过一定的等级结构折叠盘曲而形成。从遗传组成来说，两者的 DNA 和蛋白质是相同的，它是经过细胞分裂间期 DNA 和蛋白质复制而形成，在有丝分裂中期，两条染色单体通过着丝粒排列在赤道板的两侧，有丝分裂后期两条染色单体分开，并向两极移动，并最终分布于两个子细胞中。

（二）染色体功能实现的三要素

任何真核生物染色体的生物学功能都严格依赖于三种 DNA 序列结构：着丝粒、端粒和复制起点。这一点已被酵母人工染色体的成功构建所证实。

1.着丝粒在染色体有丝分裂和减数分裂过程中发挥重要作用。

着丝粒是染色体在有丝分裂和减数分裂过程中实现分离的顺式作用元件（*cis* acting elements）每个正常染色体都含有一个着丝粒，它是有丝分裂（mitosis）和减数分裂（meiosis）过程中染色体分离不可缺少的因素，缺乏着丝粒的染色体片段，不可能参与纺锤体（spindle）的形成，也不会通过细胞分裂分配到子细胞中去。

某条特定染色体着丝粒的结构和功能，由位于着丝粒处的 DNA 序列所决定。在简单的真核生物中，决定着丝粒功能的序列非常短。如在酵母细胞中，着丝粒元件（centromere element，CEN）大约有 110 bp，其中包括 96 bp 和 11 bp 的两个高度保守的侧翼元件（flanking element）以及一个约 80～90 bp 的富含 AT 的中部节段（segment）。酵母细胞中的着丝粒是可互相替换的，一个酵母染色体的着丝粒可以替换另一个染色体的着丝粒而不会引起明显的后果。然而哺乳动物着丝粒的结构更加复杂，而且具有一些特异的 DNA 序列。

2.端粒封闭了染色体的末端并且维持了染色体的稳定性

端粒是染色体末端的特化结构，包括 DNA 和蛋白质，而且对染色体末端起到隔化作用。端粒的功能一般认为有以下几点：

（1）维持染色体结构的完整性　　如果失去端粒，则染色体的末端呈现高度的不稳定性，并且使染色体末端具有与其他染色体断端融合的倾向，从而发生重组或退化。端粒结合蛋白（telomere binding proteins）能够识别端粒的 3′单链区并且能在体外保护 DNA 末端，而且它们可以提供在体内保护端粒的“帽子”（cap）结构。

（2）保证染色体末端的完整复制　　在 DNA 复制过程中，后随链的合成是不连续的，它需要一些 DNA 先导序列作为 RNA 引物的模板。因此需要一定的机制来解决线性 DNA 分子末端复制所出现的难题，细菌以环形分子来解决，而真核生物已进化出特定的 DNA 序列和特殊的酶来解决这一难题。

（3）在维持细胞核的三维结构及染色体配对中起一定作用　　这个可能的功能是在观察染色体末端定位于核膜边缘时提出的。通过与细胞膜结合，端粒可以用来锚定染色

体，因此在细胞分裂间期为染色体的重新组织提供作用力，并避免染色体臂的缠绕，另外端粒与核膜的结合，也许简化了同源染色体间的相互寻找，从而降低了染色体配对的复杂性。

3.复制起点是DNA复制起始和染色体数目的维持所必需的

正常二倍体细胞的DNA在一个细胞周期中仅仅复制一次。复制起始的控制由一些关键的顺式作用序列来决定。这些顺式作用序列可由反式作用的蛋白质结合来起作用。

真核生物复制起点在酵母细胞中已进行了大量的研究，通过遗传分析可以推测出复制起点的大致位置。在这项研究中，利用了细菌质粒来转化缺乏复制必需基因的酵母细胞，而在细菌质粒中含有宿主所缺乏的必需基因，且添加了一段酵母DNA随机序列来检测起始自动复制的能力。由于质粒中的细菌复制起点在酵母细胞中没有功能，因此少数转化效率较高的质粒中所插入的酵母DNA片段一定具有负责高效复制额外染色体的元件，称为自主复制序列(autonomously replicating sequence，*ARS*)。

*ARS*被认为是真正的复制起点。在一些研究中，已经在特殊的染色体位置上定位了*ARS*，而且证实了DNA复制在该位置起始。*ARS*仅包括50 bp的DNA序列，富含AT，有一个保守的核心共同序列和一些不完整的该序列拷贝。另外*ARS*还包括了一个转录因子结合位点和结合于起始点的蛋白质复合物。

哺乳动物起始位点由于缺陷型个体的缺乏而知之甚少。虽然对某些起始位点进行了研究，但这些研究没有鉴别出一个唯一的复制起点，这也使人们推测哺乳动物的复制可能是由多位点起始。最近，Dobbs等利用计算机对几个可能的真核生物(包括人类和其他哺乳动物)的复制起点进行了分析，鉴别出了DNA的共同序列：WAWTTDDWWWDHWG-WHMAWTT，其中W=A,T;D=A,G,T;H=A,C,T;M=A,C。

二、染色体的超微结构

(一) 染色体包装的结构模型

染色质丝的化学成分主要是：脱氧核糖核酸(DNA)、组蛋白、非组蛋白、少量的核糖核酸(RNA)及一些微量元素。DNA约占染色体组成成分的30%～40%，组蛋白含量与DNA的含量大致相等；非组蛋白的含量变化很大。在高等生物染色体中，蛋白质部分一般约占50%以上。

众所周知，组蛋白对染色体的结构有重要作用，它不但是染色体的主要组成成分，同时有证据表明，染色质丝的螺旋化和超螺旋化也是由DNA双螺旋特定部位的组蛋白所控制。那么染色体中的组蛋白与DNA是怎样结合的？是通过怎样的过程和步骤使染色质纤丝包装成染色体的？

1.核小体是染色体包装的基本单位

Kornberg(1974)通过大量的研究工作提出了染色质结构的念珠模型(beads-on-a-string)。他主张染色质的基本结构是由一系列核小体(nucleosome)相互连接而成的念

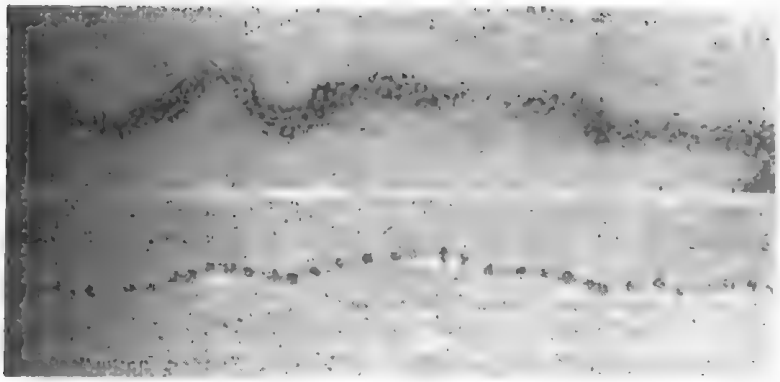


图 1-4　电镜下染色质的念珠结构(引自 P. J. Russell, 1990)

珠状。核小体是由 200 bp DNA 双螺旋与一个组蛋白八聚体和一个分子的组蛋白 H1 所组成。这一工作继而得到 D. E. Olins 和 A. O. Olins 工作的支持。Olins 夫妇用温和的方法破坏鸡红细胞核,将染色质层铺在电镜铜网上。在电镜下观察,发现染色质的基本结构是无数小珠串起来而形成,他们将这种小珠称为纽体(ν-body)。小珠直径为 10nm。核小体是由核小体核心(nucleosome core)和一条含有 H1 组蛋白的间隔(spacer)DNA[又称连接 DNA(linker DNA)]所组成。核小体的核心由组蛋白 H2A、H2B、H3、H4 各两个分子所组成的八聚体和在八聚体上缠绕 1¾圈的 146 个碱基对的核心 DNA(core DNA)所组成。若结合上 H1,则 DNA 双螺旋缠绕两圈,有 165 个碱基对。165 个碱基对加上组蛋白核心及 H1 称为染色质小体(chromotosome)。H1 与 DNA 双链的进出口端相结合,封闭了核小体核心。间隔 DNA 连接在染色质小体之间,长度约为 35 碱基对。

组蛋白是一种碱性蛋白质,等电点一般在 pH10.0 以上,是由于它们含有较多的碱性氨基酸的缘故。提取组蛋白可用 0.2mol/L H_2SO_4,组蛋白可以溶解而 DNA 却不溶解。

一般真核生物染色体上共有 5 种组蛋白,分别为 H1、H2A、H2B、H3、H4,它们在同一个体不同组织中全一样,在不同的真核生物中也很相似。其主要特征如表 1-2 所示。

在 5 种组蛋白中,除了 H1 以外,其余 4 种都有相互作用以形成聚合体的趋势,它们通过 C 端的疏水氨基酸相结合,而 N 端的碱性氨基酸则向四面伸出以便与 DNA 分子相互作用。4 种组蛋白分子中,H3 和 H4 亲和力最强,可形成四体,H2A 和 H2B 亲和力也很强,H2B 和 H4 也容易结合,H2A 和 H3 之间反应弱,H2A 和 H4 以及 H2B 和 H3 之间反应更弱,4 种组蛋白的结合方式如图 1-5。

表 1-2　组蛋白的某些特性

组蛋白	分子量(kDa)	氨基酸组成	种类的变异	每 200 bp DNA 的分子数
H1	24	富含赖氨酸	广泛	1
H2A	14	赖氨酸含量中等	有相当的保守性	2
H2B	14	赖氨酸含量中等	有限的保守性	2
H3	15	富含精氨酸	高度保守	2
H4	11	富含精氨酸	高度保守	2

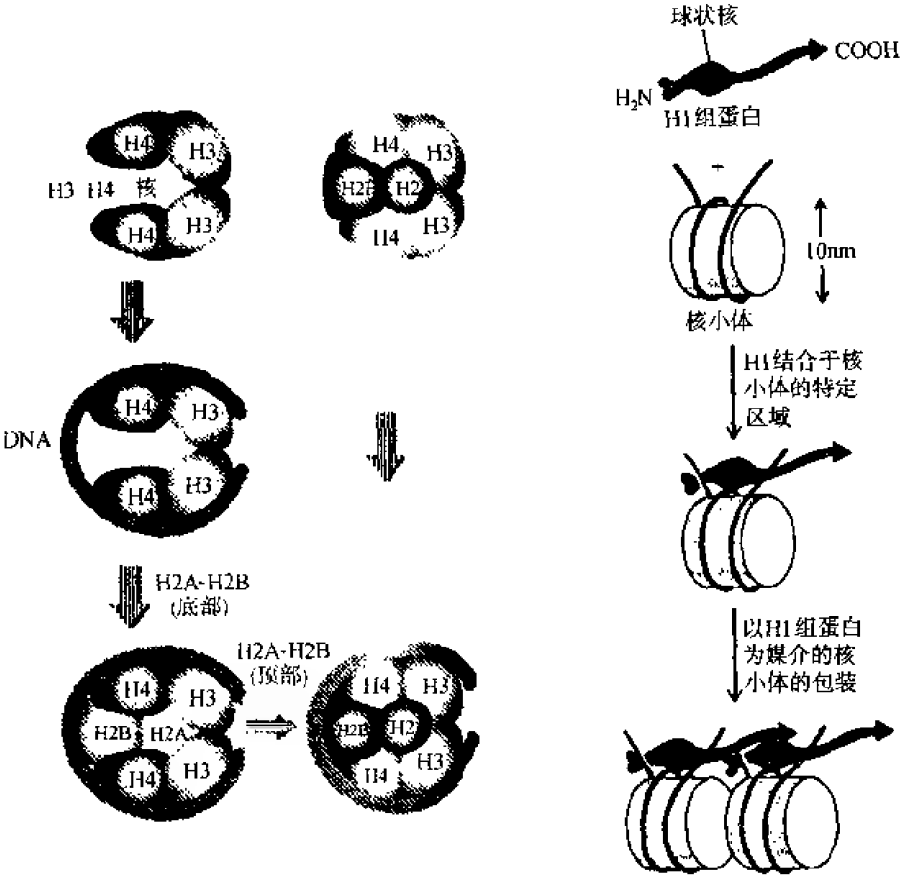


图 1-5　核小体组装及 H1 组蛋白分子与相邻核小体包装的关系

8 种组蛋白分子根据配对的亲和力差异，组成了直径 10nm，高 6nm 的扁圆柱形八聚体核心，直径 2 nm 的双链 DNA 分子在八聚体的外面缠绕 1¾圈。此种结构，构成了染色质包装的基本结构单位——核小体结构。核小体结构的组装顺序是：H3 和 H4 首先形成四聚体$(H3)_2(H4)_2$，DNA 分子在其上缠绕一圈，H2A、H2B 形成 2 个异二聚体组合于四聚体的两侧并结合约 30～40 bp 的 DNA 序列，最后组蛋白 H1 与特定的 DNA 构象的结合，封住了核小体的进口和出口(如图 1-5 所示)。这一模型与 Felsonfeld 等人实验结果相符，即在缺少 H2A 与 H2B 的情况下，H3、H4 也能与 DNA 形成类似的核小体结构，而单独的 H2A、H2B 或 H2A-H2B 则无此能力。

上述结构模型，不仅得到了电镜工作的支持，而且得到生化工作的支持。如用小球菌核酸酶消化染色质时，经过蔗糖梯度离心及琼脂糖凝胶电泳分析，发现切下的片段都是 200 bp 为单位的不同倍数的 DNA 片段，与电镜下所观察的单体、二体、三体等不同倍数的核小体是完全一致的，如图：提取电泳。实验结果表明，两个核小体之间的连接区更多地暴露于酶中，因而易从这些部位切断。如经进一步消化，则间隔 DNA 被降解，随着降解程度的不同可分为含 165 bp 的剪切核小体和 146 bp 的核心颗粒。核心上无 H1 组蛋白说明，H1 结合在间隔 DNA 上，并介于两个核小体核心之间。H1 组蛋白分子不仅维持

了核小体的结构稳定，而且可能对于形成染色体的更高级结构即染色质丝的进一步螺旋化具有重要意义。

2.染色体包装的螺线管模型

1976 年 J. T. Finch 等将鼠肝细胞核用微球菌核酸酶处理，制备出平均由 40 个核小体组成的染色质，在电镜下观察，发现在有 EDTA 存在时，可看到直径为 10 nm 的纤丝呈松弛的螺旋状态；如果有一定量的镁离子存在，则是螺距为 12～15 nm，直径为 30～50 nm 的紧密缠绕的螺旋状态。对于去掉了 H1 的染色质，则无此效应。由此可推测螺旋的形成必须有组蛋白 H1 参与。根据这些结果，Finch 等提出了一个螺线体模型(solenoid model)。

螺线体由核小体链呈螺旋形旋绕形成，其直径为 30 nm，呈中空的管状结构。螺线体的每一周螺旋包括 6 个核小体，因此 DNA 长度缩短了 6 倍。在电镜下经常观察到直径 30 nm 左右的细丝，可证明这一点。如果把核小体作为染色质的一级结构，则螺旋体就是染色质的二级结构。

虽然在一级结构和二级结构上没有争议，但从螺旋体如何进一步包装成染色体则有不同看法。由于分子生物学的发展和新方法的利用，支架环模型越来越受到人们重视。

3.染色体包装的支架环模型

典型的细胞核直径约有 5 μm(5×10^{-4} cm)，作为 DNA 二级结构的 30 nm 染色质纤维约长 0.1 cm，因此肯定还有更高级的包装序列。

Laemmli 等以酸性葡聚糖竞争性地除掉染色体上的组蛋白之后，在染色体的核心中显出蛋白质支架(scaffold)。两条染色单体的非组蛋白支架在着丝粒区域相连接，DNA 链散开围绕在支架周围，并和支架相连接，在电镜照片中可以发现无数环结构(如图 1-6)。每个染色单体均由染色质纤维所构成的环所组成，每个环大约含有 75 kb 的 DNA。环附着在中央酸性非组蛋白所组成的支架上，其中值得注意的是拓扑异构酶 II，该酶可以使 DNA 双螺旋中的两条 DNA 链解旋。拓扑异构酶 II 和其他染色质蛋白结合于富含 AT 的序列，这些几百个碱基对的富含 AT 序列(65%)所构成的支架附着区域(scaffold attachment region，SAR)，是染色质环附着的基本条件。由此而形成的支架环(scaffold-loop)复合体进一步包装形成染色单体(如图 1-7)。

上述环结构在特殊类型的染色体——灯刷染色体和多线染色体也可以观察到，在大肠杆菌染色体中(环状无组蛋白)，也可以观察到相似的环结构域(loop domain)，暗示了环也许是染色体的一般特征。

(二) 特殊染色体的支架环结构

1.灯刷染色体

上面已讨论了染色质的包装顺序，但是在间期和前期初，由于染色质纤维太细并缠绕在一起，因此，很难清楚地分辨出整条染色体的结构。然而在一些特殊类型的细胞中，如

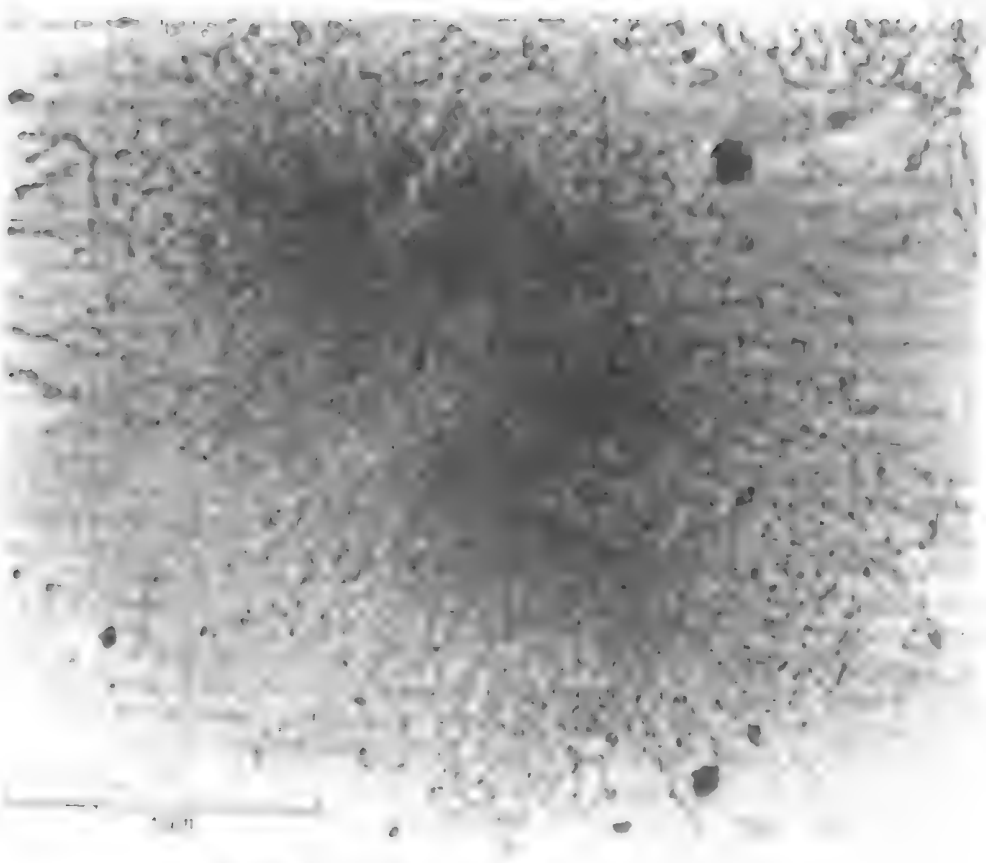


图 1-6 [illegible]

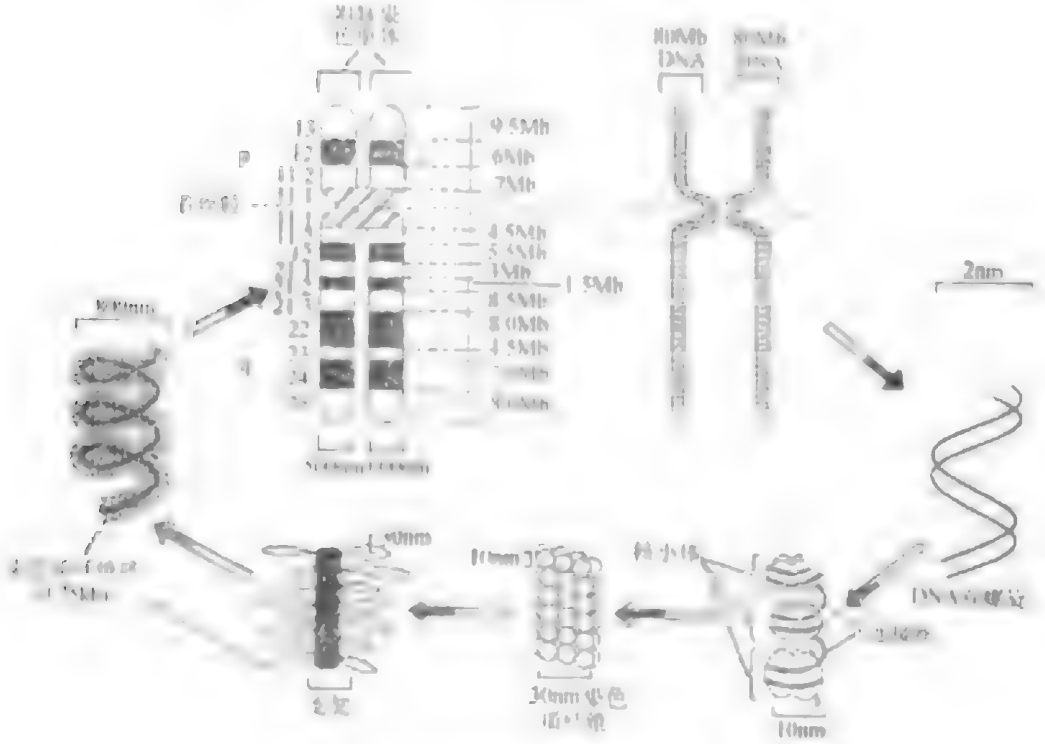


图 1-[illegible] [illegible]

在减数分裂过程染色体已配对的生长卵母细胞(未成熟卵)中,由于具有很高的 RNA 合成活性,从而形成许多伸展的染色质环结构,新合成的 RNA 与蛋白质形成 RNA-蛋白质复合物并包被在这些伸展环上。因此即使低倍镜下观察,也能清楚地看到所谓"灯刷染色体(lampbrush chromosome)"(如图 1-8)。

图 1-8 两栖类卵母细胞灯刷染色体

灯刷染色体呈典型的双线期交叉结合二价体状态(如图 1-9)。其上有排列不规则的呈串珠样的染色粒,并有成对的脱螺旋染色体环沿轴心向外伸出,许多环进行着活跃的转录,染色粒由两侧的高螺旋化的非转录区域所构成,沿染色体轴心在相邻染色粒之间存在着既无螺旋化又无转录活性的较短的染色质连接区。杂交试验表明,不同物种环的伸展方式和数目不同,而某一物种特定的环总是具有相同的序列和伸展方式,说明环的数目和形式具有种的特异性,因而可作为染色体编号的标志。

一套灯刷染色体上大约有 10 000 个环,每个环由一个转录物或几个转录物组成,转录子的 RNA 分子可能是 hnRNA,有些转录出的 RNA 分子比转录它的模板 DNA 分子短,可能在转录尚未完成之前,加工剪接过程已经开始。在卵母细胞阶段,细胞处于双线期,需表达大量 mRNA 并翻译成多种蛋白质,以满足早期胚胎发育的需要。

在灯刷染色体上,绝大部分 DNA 是以染色粒形式存在,不进行基因表达;而环的轴心是 DNA 分子。在电镜下观察每个环的轴心还伸出近乎垂直的细丝,每条细丝是一条转录出的 RNA 分子并与蛋白质组成复合体;环的粗细变化很不规则并与转录出的 RNA 长短有关。细的部分是转录起始点,RNA 分子较短,而随着 RNA 的延长,环链的直径逐渐加大。当细胞分裂进入到中期,环结构消失。

2.多线染色体

由于不同的特性,染色质上的单个区域在某些昆虫的细胞中也可以清楚地观察到。许多果蝇细胞经过多个复制周期的 DNA 合成而细胞不分裂,因此形成巨大细胞,包含了

几千倍的DNA复合物，又称多倍体，所有的纯合染色体结合在一起，形成多线染色体(polytene chromosome)。这些细胞的多倍体-多线性的转换证明这两种染色体状态具有相关性，多线染色体的基本结构与正常染色体相似。

由于体积巨大以及体细胞染色体的精确联会，延长了染色体轴心和避免了缠绕，即使在低倍镜下也可清楚地观察到多线染色体(如图 1-10)。通过对果蝇的唾液腺多线染色体研究发现，每条染色体包含了 10 个复制周期的 1024(2^{10})条 DNA 纤维。而且沿着染色体的纵长分化成相间的带纹，深者称为带(band)，浅者称间带(interband)。约 85% 的 DNA 位于带内，15% 的DNA位于间带。深染的带比间带的螺旋化程度高。通常认为每

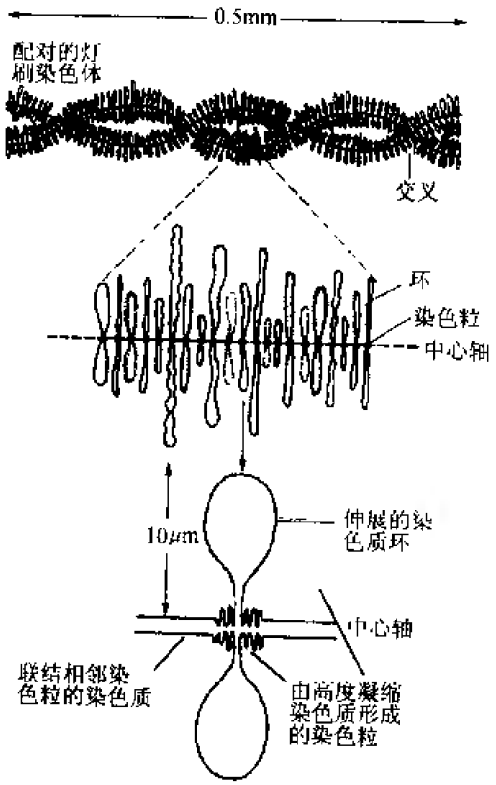


图 1-9 灯刷染色体的支架环结构

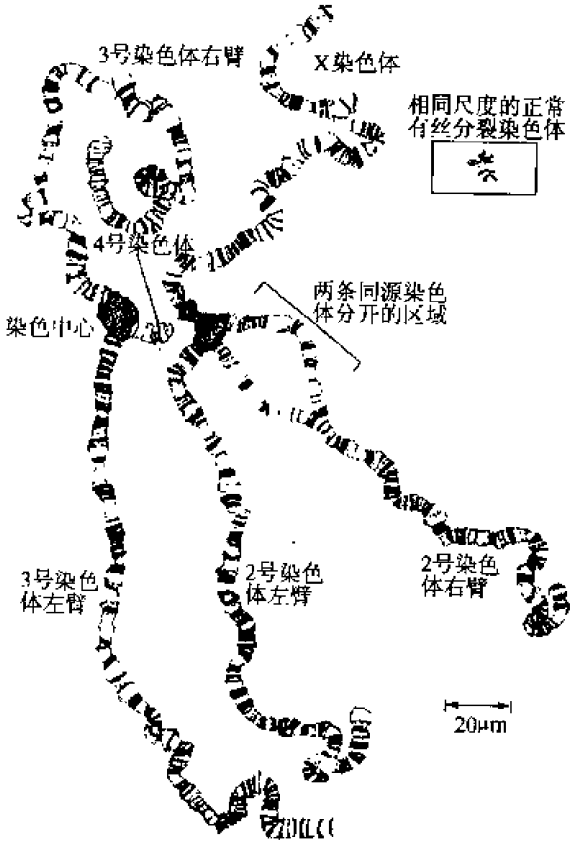


图 1-10 果蝇唾液腺细胞多线染色体

条带都代表了一个高度折叠的环结构(如图 1-11)。果蝇基因组约含有 5000 条带，根据条带和间带的宽度，估计每条染色质纤维的每条带含有 3000～300 000 bp，每条带均可识别并给予标号，从而得到多线染色体的标识图。

在染色体结构尚未清楚之前，就已注意到多线染色体的形态变化与基因表达相关。在多线染色体上常可发现局部膨大的现象，称之为胀泡(puff)，其实质就是带上的染色粒或者是折叠的环松解(如图 1-12)。二者的结构没有原则上的不同，只是胀泡部 DNA 环更为松解。胀泡的出现与基因的转录有密切关系，幼虫的发育阶段不同，则有不同的胀泡

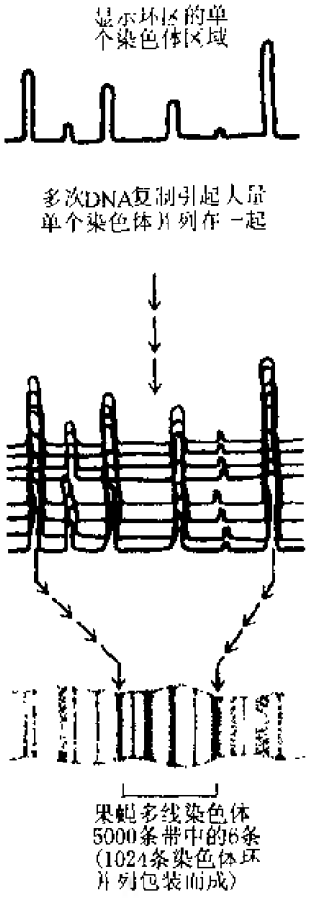


图 1-11 多线染色体上的带与环支架结构的关系

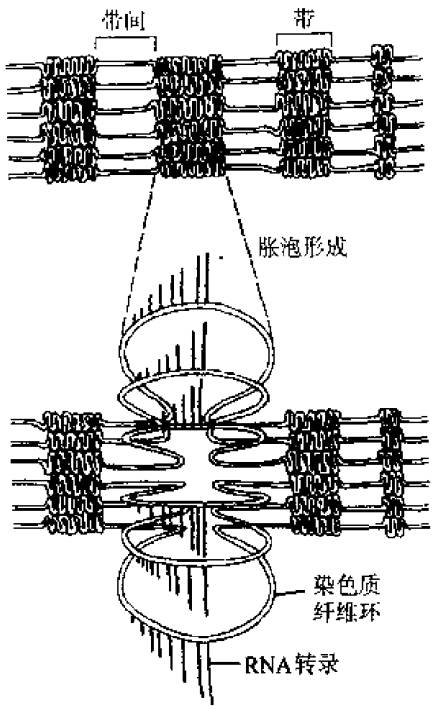


图 1-12 多线染色体的胀泡及与转录的关系

形成和消失。如用 ^{3}H尿嘧啶核苷掺入多线染色体并用放射自显影检测，发现胀泡被标记。说明胀泡部位进行着活跃的基因转录。这一点也可以用 DNA 酶 I 来证明。在果蝇基因组中由于转录部位敏感位点和超敏感位点的存在，胀泡部位往往受到 DNA 酶 I 的优先作用，因此该部位的染色质又称为活性染色质(active chromatin)。多线染色体从它的发现开始便引起人们的极大兴趣，随着分子生物学的发展，该种染色体已成为基因表达调控研究的理想材料。

(三) 常染色质和异染色质

根据染色反应，染色体中的染色质可分为常染色质(euchromatin)和异染色质(heterochromatin)两种。常染色质是指细胞分裂间期染色质丝折叠盘曲程度较小，进行细胞染色时染色较淡，随着细胞分裂的进行，这些染色质区段由于逐步的螺旋化，从而染色逐渐加深。这样的染色质区称为常染色质。而着丝粒、染色体末端部分及核仁组织区等部位的染色质丝，在细胞分裂间期就折叠盘曲得非常紧密，因而染色较深，称作染色中心(chromocenter)，随着细胞分裂的进行，其折叠盘曲程度没有太大的变化。这部分染色质

区段称为异染色质。在间期或早前期，由于异染色质的高度螺旋化，用碱性染料染色很深，这种现象称为正异固缩现象(positive heteropycnosis)。随着细胞分裂的进行，由于常染色质区段螺旋化的发展，使异染色质区段染色相对较淡，从而呈现负异固缩现象(negative heteropycnosis)。在细胞分裂不同时期呈现不同的染色反应，称为异周性(allocycle)。利用这种异固缩现象确定异染色质区所在的位置，为绘制某一物种的染色体核型图提供了有用的标记。

在遗传功能上，常染色质含有结构基因，而且具有较强的转录活性，其结构基因按孟德尔遗传方式进行遗传。异染色质则相反，它通常不含有结构基因，但具有某些遗传效应。根据其结构和行为上的差异，可将异染色质分为两类：①组成型异染色质(constitutive heterochromatin)，其DNA序列从来不转录，是一种永久性的异染色质结构，它不含或很少含有结构基因，如卫星DNA；②兼性异染色质(facultative heterochromatin，又称功能性异染色质)，它在某些情况下为常染色质，但在一定的条件下(不同性别、不同组织、不同的发育阶段)可以成为异染色质。兼性异染色质具有结构基因，但在细胞学上呈异固缩现象，遗传上是非活化型，且只出现在特定的情况下，表明它是由生理和发育的条件所决定。典型的例子就是哺乳动物雌性个体中的两条X染色体中有一条随机失活，形成巴氏小体，从而失去转录活性，维持正常的基因剂量效应，这个作用称为异染色质化。

遗传物质的异固缩现象可能是关闭基因活性的一种手段，常染色质和异染色质的相互转换，可能在结构基因的表达调控上起一定作用。

三、染色体的大小和数目

1.染色体数目

在真核生物中，每一物种都有其特定的染色体数。如人有46条，黄牛有60条，猪有38条，果蝇有8条，小鼠有40条等。虽然有时两种动物会有相同的染色体数目，如猪有38条，猫也有38条，鸡有78条，而狗也有78条。但是从染色体的形态、大小、着丝粒的位置以及基因结构和功能上却存在着很大差异。

对于同一物种来说，染色体的数目是恒定的，这对维持物种的遗传稳定性有重要的意义。绝大多数高等动物都是二倍体($2n$)，即每一体细胞中有两套同样的染色体，这两套染色体分别来自该个体的两个亲本，来自亲本的每一配子的一套染色体，称为单倍体(n)。所以，在体细胞中染色体都是成对存在的，这种一条来自父方，一条来自母方，在大小、形态、着丝粒的位置、染色粒的排列等都相同的一对染色体称为同源染色体(homologous chromosome)。在真核生物的体细胞中，都有一对性染色体(sex chromosome)。如果蝇$2n=8$，其中6条染色体雌雄个体都一样，称为常染色体(autosome)，其余两条随性别不同而不同，称为性染色体。在雌性中，两条性染色体的形态、大小、着丝粒位置均相同，称为X染色体；而在雄性中，两条性染色体只有一条与雌性X染色体相同，而另一条与X染色体存在着很大差异，称为Y染色体。

在哺乳动物中，性染色体一般属于上述类型。但在鸟类中，其性染色体在雌性中为异型，在雄性中为同型。为避免混淆，通常把鸟类的雌性异型性染色体称为ZW型，雄性为

ZZ 型。在某些昆虫中，有的雄性只有一条性染色体，也有的雌性只有一条性染色体。

在真核生物的不同物种之间，染色体数目差异很大，目前已知染色体数目最少的物种是马蛔虫(*Ascaris megalocephala*)，它只有 2 条染色体($n=1$)，而有一种蝶类(*Lysandra nivescens*)却有 380 条染色体($n=190$)。在高等植物菊科(*Haplopappus gracillis*)的染色体只有 4 条($n=2$)，而真蕨纲瓶尔小草属(*Ophioglossum*)其染色体多达 1 020 条($n=510$)。

染色体的数目多少，并不反映物种进化的程度。但是染色体的数目和形态特征对鉴别物种之间的亲缘关系，特别对近缘物种的分类常常具有重要意义。

2. 染色体的大小

除染色体数目外，还需要注意染色体大小。因为生物体的基因组 DNA 包装在染色体上，染色体的大小及数目直接反映了基因组的大小。通常以每单倍体染色体组中的 DNA 含量来表示基因组的大小，称为生物体的 *C* 值(*C*-value)。*C* 值是单倍体染色体的 DNA 总量。一个物种的 *C* 值是恒定的，不同物种之间 *C* 值差异很大，如人的基因组大小为果蝇的 20 倍，而有尾两栖类的平均基因组要比人的大 10 倍。

四、染色体分带

染色体分带(chromosome banding)技术的开创，是细胞遗传学、分子遗传学、细胞工程学研究中的一大突破。1968 年，Casopersson 首次报道了用荧光染料喹吖因技术处理的染色体，在紫外光激发下染色体臂显示出不同的带纹。同一研究组继续用喹吖因及其他相关的化合物对各种生物包括人类的染色体分带进行了研究，进而推进了染色体带型研究技术的发展。如 G 带、C 带、R 带、T 带等，这些分带的研究对于核型分析、染色体鉴别以及分类学上都有重要的意义。

1. 荧光分带法

1969 年，Caspersson 用喹吖因(quinacrine)对处于有丝分裂中期的染色体染色，在紫外光激发下，在荧光显微镜下观察，发现每对染色体上出现了各具特色的明暗相间的带型，命名为荧光分带法(Q-banding，Q 带)。在对小鼠染色体研究方面，由于 Q 带技术的发展，取得了显著的进展。小鼠的染色体在未采用显带技术以前是很难识别的，现在根据 Q 带带型，已经建立起小鼠染色体命名系统。喹吖因易与 DNA 富含 AT 部位结合，Q 带可能就是染色体富含 AT 的部位。Q 带技术分类简便，可显示独特带型，但标本易褪色。

2. 吉姆萨分带法

1977 年，Caspersson 发现用稀释的胰蛋白酶溶液、尿素或蛋白酶预先处理有丝分裂中期染色体，然后用 Giemsa 染色，在光学显微镜下观察，可看到染色体上出现染色深浅不同的带纹，称为吉姆萨分带法(G-banding，G 带)。G 带与 Q 带十分相似，而且方法简便，费用低廉，因而应用十分广泛。G 带技术在脊椎动物中所染出的带型，每条染色体都不同，因而可以明确地鉴别每一条染色体。

G带的性质是一个争议较多的问题，有人认为G带反映了染色质的螺旋化程度，有人解释为是由于染色体的组蛋白和其他蛋白质沿染色体成簇分布的结果，也有人认为染色体的DNA和蛋白质都参与了G带程序所发生的细胞化学反应过程。

3.反带法

用高温处理染色体标本，用低pH值(pH 4.0～4.5)的磷酸盐溶液，在88℃条件下处理11分钟，然后以Giemsa染色，或者将染色体老化7～10天，经磷酸缓冲液处理，必要时在85℃温育20～25分钟，以吖啶橙染色，其结果所显示的带和G带的明暗相间正好相反，因而对G带端部不着染部分的可用此种反带(reverse banding，Riband ing)法，又称R带法，从而将末端清楚地显示出来。两种方法并用可确定染色体的丢失部位。

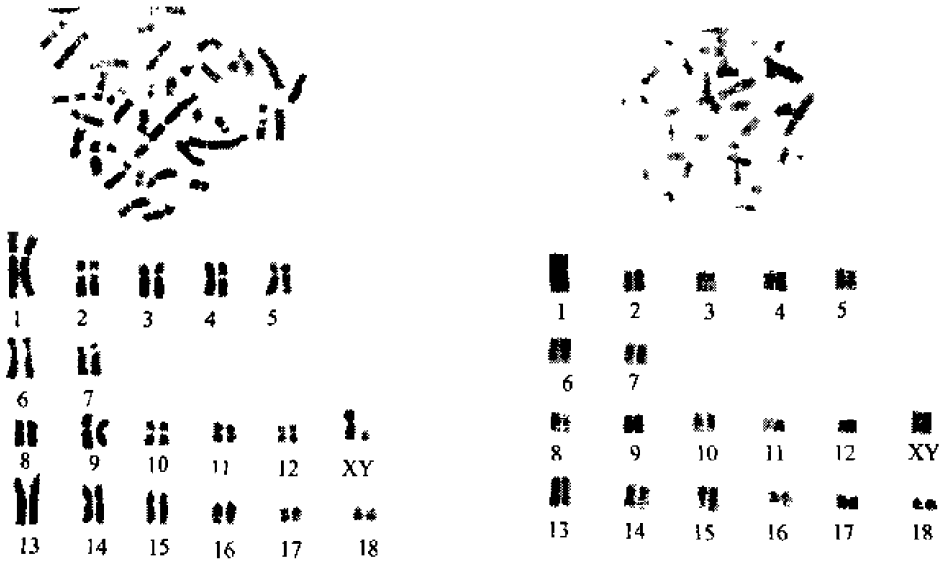


图1－13　家猪染色体分带组型图
左：G带(公)；右：C带(母)。

4.C分带法

这种带型能显示出染色体中组成型异染色质的区段，故称C分带(C-banding，C带)法。组成型异染色质区段在染色体上的含量较为稳定。主要位于着丝粒、端粒附近和核仁形成区，以及Y染色体的长臂上。在间期和有丝分裂早期，当常染色质处于伸长状态时，它处于异固缩状态，因而易被染料着色，但在中期时，反而不易着色。因此要使异染色质区段成为在显微镜下可识别的区段，需采用特殊方法进行处理。这些方法包括用酸、碱、盐或高温处理中期染色体，使DNA发生变性，然后在一定条件下缓慢复性。异染色质区段由于存在高度重复的DNA序列，因此可以复性，而常染色质区段的低重复序列和单一序列DNA不复性，从而产生不同的染色反应。

C带技术对辩认哺乳动物的Y染色体是非常有用的。因为Y染色体的异染色质比例较大，采用C分带法，染色效果尤其明显。C带技术也用于分析植物染色体，而且植物的异染色质显现比动物的更为突出。

5.N 分带法

又称 Ag-NOR 染色法，是显示核仁组织区(NOR)的显带技术，其主要试剂是甲酸-硝酸银混合液，所以也统称为硝酸银染色技术。该技术能使有活性的核仁组织区特异染色，呈黑色。现已知道，编码 rRNA 的 rDNA 基因位于 NORs 上，但只有具有转录活性的 rDNA 基因，或者已经转录过，并在其上仍有残余的与 rRNA 相联系的蛋白质的 NOR 才被银染呈黑色，故其着色频率与细胞中具有转录活性的rDNA基因相一致。因此，在不同生理、病理条件下，计数细胞中银染的 NOR 频率，以分析有活性的 rDNA 基因的动态变化，是目前探讨基因功能的方法之一。此外，人类近端着丝粒染色体的随体间常发生联合，这种联合可能是造成近端着丝粒不分离和易位的原因。应用银染技术，可在发生联合的染色体间看到银染物相连，因此用它作为近端着丝粒染色体联合的主观标准，比客观判断要准确可靠。所以银染技术已逐渐受到重视并开始应用于细胞遗传学研究，在肿瘤细胞遗传学等方面应用日趋广泛。技术本身也在不断改进和发展。

6.高分辨显带法

1979 年，Yunis 等应用氨甲蝶呤处理培养中的细胞，使其同步化，然后用秋水仙胺短时间处理，使其出现大量晚前期和早中期的分裂相。其中早中期染色体可出现 555～1 256 条带，而过去在中期染色体上最多只出现 320～554 条。在这样的标本上，用光镜就可观察更细微的染色体异常和对有结构重排的染色体断点作精确定位。此检测法称为高分辨显带法(high-resolution banding)。目前，高分辨率研究又有新进展，用放线菌素 D 作用于 DNA 合成后期(G_2 期)细胞以阻碍染色体浓缩时特殊蛋白与染色体结合，从而使染色体变得更细长，所显带纹可达 5000～10 000 条，使分辨率进一步提高。一般认为，人类核基因组大约有 65 000～80 000 个基因。故高分辨技术不仅对染色体变异进行精确定位，而且在基因定位和基因图谱的详细绘制也有重要意义。

7.T 分带法

T 分带法(terminal banding)是对染色体末端区的特殊显带技术，其带型在分析染色体末端的缺失、添加、末端易位及断裂点的定位上有一定价值。

8.BG 分带法

观察姐妹染色单体互换(sister chromatid exchange，SCE)，过去用放射自显影的方法。后来使用 5-溴脱氧尿嘧啶核苷(5-bromodeoxyuridine，BUdR)来替代 3′H-TdR，从而建立了一种简易的显示姐妹染色单体分化染色的方法。这种方法技术操作容易，分辨率高，对分析染色单体互换和结构异常都有价值。其主要机制是在 DNA 复制过程中，BUdR 可代替胸腺嘧啶核苷酸(T)的位置，掺入新合成的核苷酸链中。当细胞在含有 BUdR 的培养液中培养一个分裂周期后，两条姐妹染色单体的 DNA 双链化学组成就有差异，其中一条 DNA 链的 T 全部被 BUdR 取代，而另一条链未发生变化。当细胞再在含 BUdR 的培养液中经历一个周期的培养时，则一条染色单体中只有一条 DNA 链含有 BUdR，而另一条染色单体的两条 DNA 链都含有 BUdR，由于有 BUdR 的双链，其螺旋化程度降低，使染色

单体与染料的亲和力下降，因此当用荧光染料染色时，会出现不同强度的染色反应。其中，两条 DNA 链均含有 BUdR 的染色单体发生的荧光较弱，而只有一股 DNA 链含有 BUdR 的染色单体的荧光较强。利用这个特点可清楚地看到姐妹染色单体的交换情况。SCE 检测方法即可用于培养细胞，也可用于小动物（小鼠、鱼等），来检测环境中致癌、致畸、致突变物质的生物学效应，也可用于细胞周期的分析。

五、核型与核型分析

1. 核型与核型分析

核型（karyotype）代表了一个个体、一个种甚至一个属或更大类群生物的特征，动物、植物和真菌都有自己特异的核型。核型分析对于研究生物的遗传变异、生物的系统演化、物种之间的亲缘关系、某些物种的起源、远缘杂交及染色体工程、辐射遗传效应、人类遗传性疾病等都有重要意义，而且也推动了生物工程研究的发展。

核型（或染色体组型），是指一个物种所特有的染色体数目和每一条染色体所特有的形态特征（染色体的长度、着丝粒的位置、臂比值、随体的有无、次缢痕的数目及位置、异染色质的分布等）。核型是物种最稳定的性状和标志，通常在体细胞有丝分裂中期时进行核型的分析鉴定。核型分析（或染色体组型分析）是指利用显微摄影的方法，把生物体细胞内整套染色体拍摄下来，按照它们相对恒定的特征排列起来，并进行分析的过程。

一般，染色体组型分析有以下 4 种方法：

（1）*常规的形态分析*　为辨别各对染色体，需用其特有的形态特征来作为形态分析指标。如测量染色体长度，确定着丝粒位置，次缢痕及随体的有无等，也可以用臂比值、着丝粒指数（centromere index）、相对长度（relative length）来作为指标。

臂比值：染色体的长臂长除以短臂长，其数据总大于 1，$A = p/q$。

着丝粒指数：每条染色体的短臂长乘以 100，除以该条染色体的全长。

相对长度：一条染色体的全长乘以 100，再除以单倍体组全部染色体的全长，包括被测量的一条，以百分比表示。

（2）*带型分析*　在染色体组型分析中越来越多地应用分带技术。根据染色体带型的差异，可以更细微更可靠地识别染色体的个体性（图 1－14）。

（3）*着色区段分析*　染色体经低温、KCl 和酶解、HCl 或 HCl 与醋酸混合液等处理后制片能使染色体出现异固缩反应，从而产生不同的着色区段，这种着色区具有特殊性，且较为稳定。在同源染色体之间着色区基本相同，而在非同源染色体之间则有差别，从而作为染色体组型分析的一种方法。

（4）*定量细胞化学方法*　即根据细胞核、染色体组或每一条染色体的 DNA 含量及其他化学特性去鉴定染色体。由于染色体大小不同，因而 DNA 含量也不同。将染色体进行细胞化学处理和定量染色后，经光度扫描，把染色结果转化为光密度结果，从而综合成不同的参数关系，来描述染色体的微细结构、大小的差异，进而区分染色体。

在所测量的细胞中，选取一个最清晰的标准分裂中期细胞，将其染色体放大、洗印成相片，按上述各种分析指标，把成对的同源染色体剪贴在白纸板上，并附上分裂中期相片，

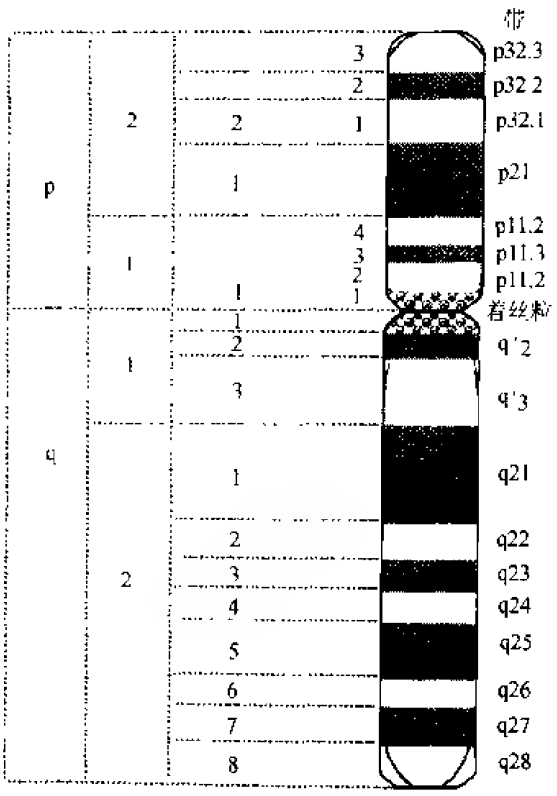


图 1-14 以分带技术显示的人类 X 染色体分带模式图

再进行翻拍而形成的图型称为染色体组型图(核型图,karyogram)。将核型图绘制成为模式图,就得到染色体组型模式图(idiogram)。

2.带型与染色体结构及基因密度的关系

浓淡相间的带型可能反映了基因组划分成异位带(isochores)的情况,这些异位带的形成是由于 DNA 片段(>300kb)的碱基构成或支架附着区的空间排布差异所致。

一般认为 G 带是复制较晚的高度浓缩的染色质,R 带为 S 期较早复制的浓缩程度较低的染色质结构。复制较晚高度浓缩的 G 带区段转录活性相对不活跃,而活性基因主要集中在 R 带。关于带型差异的结构基础,G 带和 R 带分别富含 AT 和 GC;产生 G 带的喹吖因专一性结合在富含 AT 的 DNA 序列,而通过应用专一结合 GC 区段的色霉素可以清晰地分辨出 R 带。但最近的研究表明,G/Q 带 DNA 的 AT 含量仅略高于 R 带。

说明支架环模型的新染色技术表明,富含 AT 的 SAR 沿染色单体的纵向差异分布,高度包装的 SAR 区段与 G/Q 带相联系,而未折叠的 SAR 与 R 带相联系。在这个模型中,吉姆萨可以选择性地对环染色,而根据环体的选择性染色建立了 R 带模型。

染色体区段间及整条染色体间基因密度差异很大。异染色质区域一般认为大部分由非编码的重复 DNA 序列构成,着丝粒及 Y 染色体的大部分区段也无基因存在。通过利用 CpG 岛(与 56%以上的基因有联系)探针对人类基因组杂交及与中期染色体杂交,获得了沿不同染色体纵长的基因分布,即在其近端区域基因密度较高,在某些染色体(如 19 号和 22 号)基因含量丰富而其他染色体(如 4 号和 18 号)基因贫乏。

第四节 细胞分裂[79-86]

在所有生物全部的生命活动中,繁殖后代并且保证该物种的遗传稳定性,是生命得以延续的一个重要特征。亲代将遗传物质传给后代,后代通过自身的生长发育,使其性状得以表达,从而产生与亲本相似的复制品。在这一系列过程中,细胞的分裂和繁殖是一切生命活动的前提条件。

就单细胞原核生物来说,细胞分裂就意味着无性繁殖,从而使原来的个体形成一个新个体。在真核生物中,无论是单细胞的生物个体,还是多细胞的生物个体,细胞繁殖的基

本形式是有丝分裂，通过有丝分裂，实现细胞的分化和组织的发生，从而完成个体发育的整个过程。因此，有丝分裂是真核生物繁殖的基础。

一、无丝分裂

无丝分裂，是原核生物细胞分裂方式，真核生物的某些组织和细胞也采用这种分裂方式，如肿瘤细胞和伤愈组织。其分裂方式较简单，在分裂过程中，不出现纺锤体，所以又称直接分裂。无丝分裂的分裂过程仅包括染色体（DNA 分子）的复制和细胞直接分裂两个过程。例如在细菌细胞分裂时，其环状 DNA 附着在细胞膜特定的嵴状突起上，当 DNA 复制完成后，两个 DNA 分子随着细胞膜嵴状突起的延伸而分开，在两个 DNA 分子完全分开后，细胞中间逐渐凹陷，最终分裂为两个子细胞。

二、细胞周期

细胞周期（cell cycle）是指一次细胞分裂结束到下一次细胞分裂结束所经历的整个过程。在这一过程中，细胞内的遗传物质经过复制，各种组分加倍，然后平均分配到两个子细胞中。

细胞周期又可以分为间期（interphase）和有丝分裂期（M phase）。从一次有丝分裂结束到下一次有丝分裂开始的时期称为间期。在这个时期，光镜下看不到细胞明显的变化，但是在细胞中，却进行着一系列的生化反应，包括 DNA 复制、RNA 转录和大量蛋白质的合成。间期以 DNA 合成为标志，可以分为 G_1期、S 期和 G_2期。与光镜下可见的有丝分裂相对应真正分裂的时期称为分裂期。经过分裂期，加倍的染色体平均分配到两个子细胞中，其他细胞组分也随细胞质分配到子细胞中，而最终完成整个细胞分裂过程。在该期，几乎所有的细胞内合成活动都处于停止状态。细胞在细胞周期中按 G_1-S-G_2-M 顺序而完成其增殖过程。在这个循环过程中，不同生物细胞的细胞周期持续时间差异很大，同种生物不同组织之间也不相同。对大部分动物体细胞来说，细胞周期持续时间为 18～24 小时。具体各期持续时间如图 1－15 所示。

三、有丝分裂

（一）间　　期

在细胞周期中，间期（interphase）的时间间隔较长，其原因是，在细胞周期的间期，在进入下次分裂过程之前，必须进行一系列的准备。染色质的复制是间期的重要准备工作，因为只有染色质进行了复制，有丝分裂时子细胞才能得到整套的染色体。而染色质的复制包括 DNA 复制、染色质蛋白的合成及核小体的自装配等过程，所以间期又称为合成阶段。

1. G_1期

G_1期，又称 DNA 合成前期（pre-synthetic gap phase）。在 G_1期，染色体并未复制，因

此，可认为它是 DNA 合成的一个准备阶段。在这一时期物质代谢活跃，细胞体积增大，进行着 RNA 和蛋白质的合成。在调节细胞生长中，G_1 期与 DNA 合成启动和细胞增殖调控有关。这是因为在新生组织中，连续增殖的的细胞其增殖率受 G_1 期细胞的平均时间长短所控制，细胞周期持续时间，也取决于 G_1 期的持续时间，而 S 期(S phase)和 G_2 期的时间则相对稳定。有些细胞经常从细胞周期中退出而进入一种与 G_1 相似，但又截然不同的 G_0 期。这些细胞可通过刺激离开 G_0 并重新进入细胞周期。故 G_1 期可能存在着一个限制点(restriction point)(在酵母中称为 Start)，来接受一系列特异或非特异的环境信号，使处于不利条件下的细胞停止在 G_1 期(图 1-16)，或退入 G_0 期(图 1-15)。

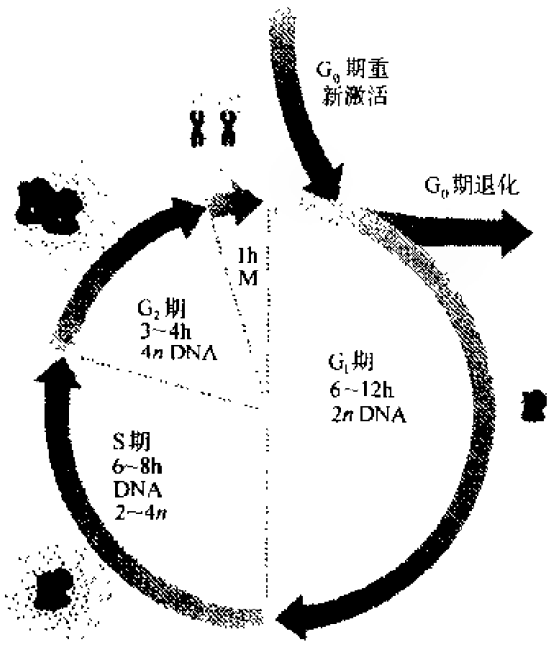


图 1-15　有丝分裂细胞周期示意图

在 G_1 期，最主要的生化反应是蛋白质和 RNA 的合成，从而为进入 DNA 合

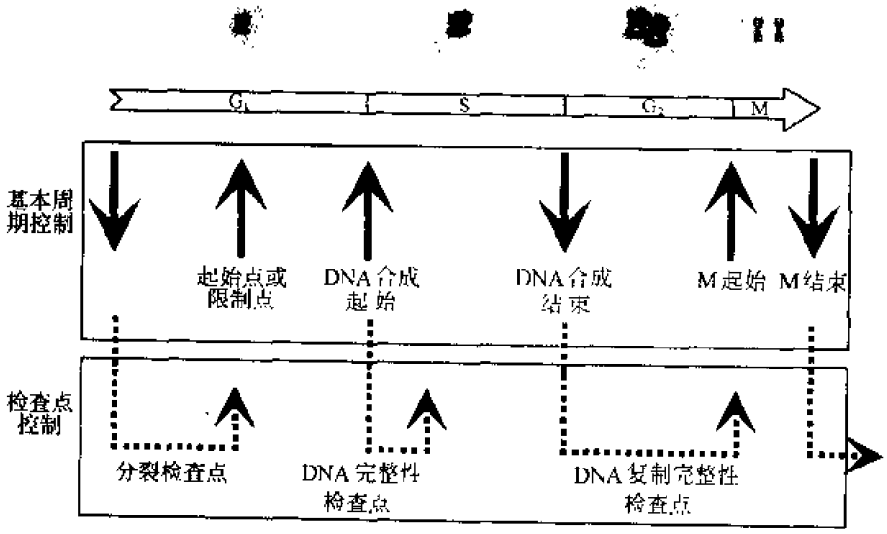


图 1-16　细胞分裂及其位点

成期准备必要的条件(如酶等)。如果主要的蛋白质和 DNA 合成被抑制，则细胞不能进入增殖周期，也不能从 G_1 期进入 S 期。另外在 G_1 期，还存在染色体构型的变化，即染色体的凝集-解凝集过程，该过程也是 G_1 期的主要反应。已知 M 期染色质经螺旋化而凝集为染色体，因此染色体解凝集是进入 S 期进行 DNA 合成的先决条件。

2.S 期

S 期(G_2 phase),又称 DNA 合成期。在该期,主要完成 DNA 的合成和组蛋白、非组蛋白等染色体蛋白质的合成,并装配成具有一定结构的染色质。

不少研究者从核质互作的角度来探讨 DNA 合成的起因。Rao 和 Johnson 等用细胞融合方法研究 DNA 合成中核质间的相互关系,他们用 G_1期 HeLa 细胞与 S 期 HeLa 细胞融合,明显观察到 G_1/S 期融合细胞中,G_1期细胞核中 DNA 合成比率增加,说明在 S 期细胞质中存在某种因素负责 DNA 合成启动,并可以作用于 G_1期细胞核中诱导 DNA 复制过程。但在所建立的 S/G_2期融合细胞中,并不能诱导 G_2期细胞核中的 DNA 合成。通过进一步实验表明,这种存在于 S 期细胞质中的 DNA 合成诱导因素是在 G_1期逐渐积累起来的。在此基础上,Rao 等提出了解释模型:当 G_1期向 S 期过渡时,DNA 合成诱导因素逐渐累积,当达到临界水平时,DNA 复制被启动,至早 S 期或中 S 期达到峰值,然后逐渐减少,在 S/G_2期边界处诱导因素量降到临界水平以下,DNA 合成停止。

在真核生物中,DNA 的复制和蛋白质的合成是相对应的。随着 S 期 DNA 在核内的合成,也进行着组蛋白和非组蛋白等染色体蛋白的合成,并迅速地运输到核内,参与新生姐妹染色单体的组装。因此,DNA 复制和组蛋白的合成有着密切的关系。当 S 期细胞被 DNA 合成抑制剂如羟基脲(hydroxyurea)作用时,组蛋白合成很快停止,其 mRNA 水平也迅速降低。DNA 合成与组蛋白合成之间关系反映了一种反馈机制,来调节游离组蛋白水平,保证新合成的组蛋白数量与新合成的 DNA 数量相对应。

S 期 DNA 的复制有一定的顺序。在 S 期早复制的部位,在下一次细胞周期中也早复制。一般来说常染色质早复制,异染色质后复制,GC 含量高的部位先复制,AT 含量高的后复制。

3.G_2期

G_2期(G_2 phase),又称 DNA 合成后期。通过辐射实验证实,若在 G_1期辐射,染色体发生畸变;若在 G_2期辐射,则染色单体发生畸变,说明 DNA 复制在 S 期完成后,细胞才进入 G_2期。G_2期细胞合成一定的蛋白质和 RNA 分子,这是细胞进入有丝分裂所必需的,如微管蛋白等。如果用嘌呤霉素、环己亚胺等蛋白质合成抑制剂作用于 G_2期细胞,则这些细胞不能进入 M 期。另外还有一些基因其产物也是细胞通过 G_2期进入 M 期所必需的。如在 ts85(小鼠 FM3A)细胞中,由于 H1 组蛋白磷酸化缺陷,从而阻止了染色体的凝集,导致细胞周期停止在 G_2期。

4.有丝分裂的其他准备

除染色质复制外,细胞准备有丝分裂的现象是细胞生长。其表现是,间期结束后分裂期之前,细胞的体积几乎增大一倍。说明进行有丝分裂的细胞不但要准备核分裂(karyokinesis),还要为细胞质分裂(cytokinesis)作准备。有丝分裂的另一项准备工作是中心粒的复制和向两极移动。

在间期的 G_1期,通常可观察到两个中心粒,它们一般在 S 期复制。复制的新中心粒称前中心粒,它比母中心粒小并与母中心粒呈大于 90°的 L 形状。前中心粒的大小逐渐

增加，直至与母中心粒大小相同。G_2期结束时完成中心粒的复制。在这时，两对中心粒仍处于核膜一边。前期开始，一对中心粒与另一对分开并沿核外缘向对侧移动，而另一对中心粒位置不动。同时在两对分开的中心粒之间出现小的纺锤丝片段。在中期两对中心粒分别定位于细胞的两极，并在两者之间形成完整的纺锤丝体系。

（二）有丝分裂期

有丝分裂期是一个连续过程，根据形态人为地划分为前期、早中期、中期、后期和末期（图 1－17）。

1. 前期

前期（prophase）发生的主要变化是染色质丝不断螺旋化而逐渐凝集成染色体，核膜核仁消失和形成纺锤体等。

标志前期开始的第一个特征是染色质不断凝集，实质是染色质的螺旋化、折叠和盘曲的包装过程。每一早前期染色体都是由两条相互缠绕的两条染色单体的染色质丝组成。它是间期的 S 期染色体复制的结果。随着螺旋化和折叠盘曲的不断进行，相互缠绕的两条染色单体逐渐分开并处于平行状态。在前期中，染色体纤维进一步缩短、变粗，可以看到两条染色单体在主缢痕处连接。两对中心体和主缢痕处的着丝粒各作为微管组织中心，分别伸出纺锤丝并逐渐形成纺锤体。在前期末，核仁消失，核膜崩解。染色体遍布整个细胞，成为它们向两极移动分离成染色单体的最好时机。

2. 早中期

核膜破裂，则细胞进入早中期（prometaphase），该期的特征是染色体剧烈活动。两极中心粒伸出的纺锤丝分别与染色体两侧的着丝粒结合形成染色体牵丝，两侧相反方向的牵引力量达到平衡，使染色体排列在中期的赤道板（equatorial plane）上，从而完成染色体在赤道板的集合（chromosome congression）和着丝粒的定向（centromere orientation）。早中期在哺乳类细胞中持续 10～20 分钟。

着丝粒完成定向后，其相应的染色体大致均匀地分布在赤道板上。然而，染色体在赤道板上的分布并不是完全随机的，可能具有某种程度的特定排列。如许多昆虫的中期染色体，较大的染色体位于赤道板的两端，而较小的染色体位于中间。与其相反，人类细胞中最大的 1 号、2 号染色体靠近中间，而较小的 Y、13、17、18 和 21 号染色体位于外缘。因此，有关染色体在赤道板上的排列机制中，可能还包括其他作用因素。

3. 中期

中期（metaphase）染色体高度螺旋化，因而比其他任何时期都要短而粗，所以适于细胞学研究，是核型分析的最佳时期。此时，两条染色体不再相互缠绕，而处于平行排列状态。染色体的两条染色单体只在未分裂的着丝粒处相互连接。中期的结束是以着丝粒的断裂和所有组妹染色单体在着丝粒处分开为标志。

4.后期

后期(anaphase)是活跃和迅速运动的时期,在这一时期中,组妹染色单体分开并向两极移动。当子染色体到达两极时,此期结束。

染色单体的分开从着丝粒处开始。着丝粒部位的分离,打破了力的平衡,在染色体牵丝的牵引下产生染色体的移动。移动时着丝粒部位在前,两臂拖后,视着丝粒在染色体上的位置不同,而呈V、L或J型。在同一细胞内无论染色体大小,它们都以相同的速度移向两极,其移动是同步的。

在纺锤体中有两种类型的纺锤丝,一种是连续性纺锤丝,它从纺锤体的一极延伸到另一极;另一种是染色体牵丝,它与染色单体的着丝粒相连接。有关染色体移动的机制有多种,有的认为是牵丝的收缩,有的认为是牵丝的收缩和连续性纺锤丝伸长的共同结果,但都缺乏足够的证据。

5.末期

末期(telophase)是子染色体到达两极后至形成两个新细胞为止,主要过程是子核形成和胞质分裂。

末期时,到达两极的两组染色体重新由核膜包围,并形成两个子核,同时染色体开始解螺旋,而丧失致密度和着色能力,染色质分散在核中。在特定的染色体上的核仁组织区部位出现新的核仁。然而在有些生物细胞中,有的核仁组织区没有活性,因此往往核仁数比核仁组织区少。

虽然核分裂和胞质分裂是相继发生的,但是属于两个分裂过程。动物细胞的胞质分裂是在细胞质的周围边缘有一个由微丝组成的收缩环(contractile ring),它的紧缩使细胞产生缢缩,最后在缢宿处凹陷使细胞一分为二。末期结束了有丝分裂周期,然后进入了新细胞周期的间期阶段。

四、减数分裂

有丝分裂是体细胞的分裂方式,其特点是分裂后的两个子细胞染色体数目相等,且与其母细胞的染色体数目相同。但是,如图1-17所示在配子(精子、卵子)形成过程中,还有另一种细胞分裂方式——减数分裂(meiosis)。

(一)减数分裂的概念

有性生殖是生物长期进化过程中所形成的更为进步的一种繁殖方式。通过受精作用,雌雄配子融合形成合子,其结果既保证了遗传的稳定性,又增加了许多新的变异,增强了生物对环境的适应能力。显然,雌雄配子在互相融合形成合子时,其染色体数目要增加一倍,因此,需要有一种机制来使配子形成过程中,染色体数目减半,从而保证合子的染色体数目和亲代相同。减数分裂是有性生殖个体性母细胞(精母细胞、卵母细胞)成熟后,配子形成过程中所发生的特殊方式的有丝分裂,它只经过一次染色体复制,却经过两次连续

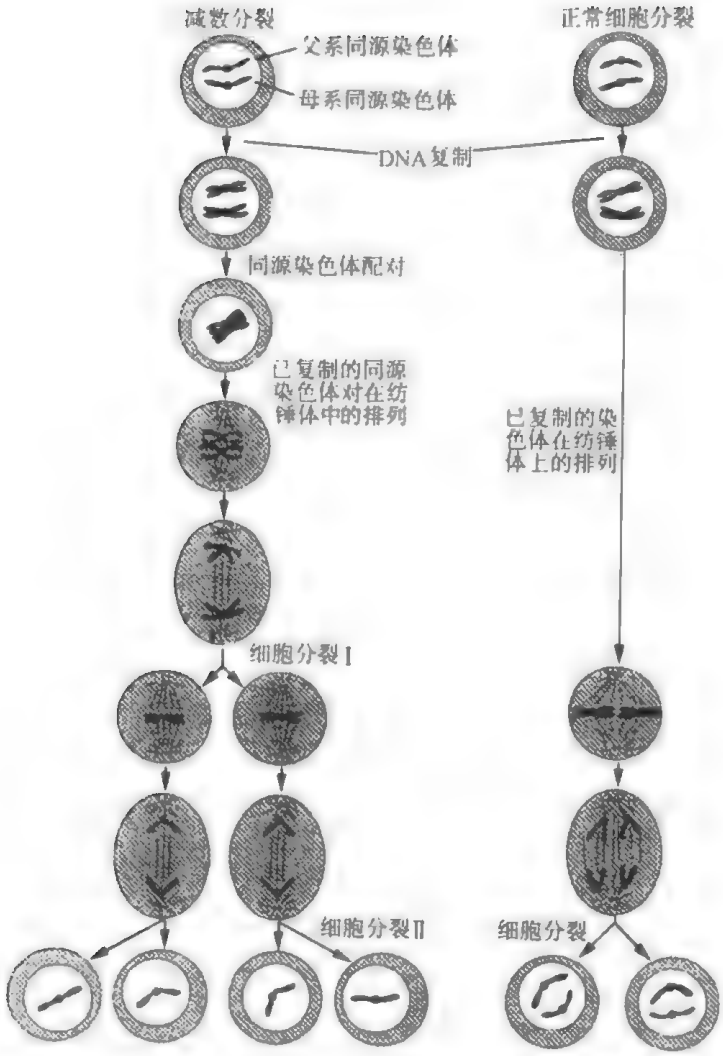


图 1-17　有丝分裂与减数分裂示意图

的核分裂，从而使子细胞的染色体数目，只有原来母细胞的 一半。减数分裂形成的子细胞，经过配子发生过程发育成配子。在染色体行为方面，可以发生联会(synapsis)和交换(crossing over)，并导致重组的发生。减数分裂过程中的重组有两种类型：一种是由核内各对同源染色体的独立分配和非同源染色体之间的自由组合而产生的染色体之间的重组，另一种是由于交换面产生的染色体内基因重组。重组的发生对于生物进化和动物育种有重大意义。

(二) 减数分裂前的间期

减数分裂前的间期与有丝分裂的间期很相似，也分为 G_1期、S 和 G_2期。不同的是减数分裂前的 S 期比有丝分裂的 S 期要长。此种延长并非由于复制活动减慢，而是由于每单位长度 DNA 复制单元的启动减少所致。另一令人瞩目的现象是在有些生物中，减数

分裂S期合成全部染色体DNA的99.7%，其余的0.3%在偶线期合成。

由减数分裂的G_2期细胞进入两次有序的减数分裂。第一次减数分裂可分为前期I、中期I、后期I、末期I。第二次减数分裂可分为前期II、中期II、后期II、末期II。两次分裂之间的分裂间期或长或短，但无DNA合成。两次分裂中以前期I最为复杂，经历时间最长，在遗传上意义也最大，可细分为细线期、偶线期、粗线期、双线期和终变期。这些时期的划分是对一个连续过程进行的人为划分。整个过程如图1-17所示。

（三）减数分裂I

1.前期I

前期I的一个重要特征是细胞核明显增大，动植物减数分裂的前期核体积要比有丝分裂前期核大3～4倍。在这一时期，染色体还表现出一些特殊行为：染色体配对、单体交换、相斥和端化作用。

（1）细线期（leptonema）　这是减数分裂过程开始的时期。在这一时期中，细胞核的体积增大，核仁也较大，染色体开始凝缩与螺旋化，呈现出类似于有丝分裂早前期的线性结构。此期线状染色体往往缠在一起，并偏于核的一侧，因此又称凝线期（synizesis）。在动物减数分裂中，虽然此期同源染色体尚未配对，但同源染色体的一端或两端附于核膜上，而且彼此靠近。因而人们推测，染色体配对可能从这里开始。有的生物在细线期细胞中，所有染色体的一端都集中在中心体所在位置的核膜上，形成花束状，所以这一时期又称花束期（bouquet stage）。花束的形成可能有助于偶线期同源染色体的联会。

（2）偶线期（zygonema）　偶线期主要是同源染色体的配对时期，各同源染色体两两配对，这一现象称为联会（synapsis）。这种配对可以想象为沿着染色体的任一点甚至若干接触点上开始，并扩展到所有的同源区段，使同源染色体像拉链一样并排排列处于配对的平衡状态。然而，有的同源染色体的配对往往是不完全的，如XY染色体既有同源区段也有非同源区段。但在通常情况下，同源染色体之间的配对是精确而专一的。在该期有0.3%的DNA合成，从构型看为Z-DNA。有人认为Z-DNA的合成可能和配对机制有关。联会所形成的特殊结构沿同源染色体纵轴分布，称为联会复合体（synaptonemal complex，SC）（图1-18）。

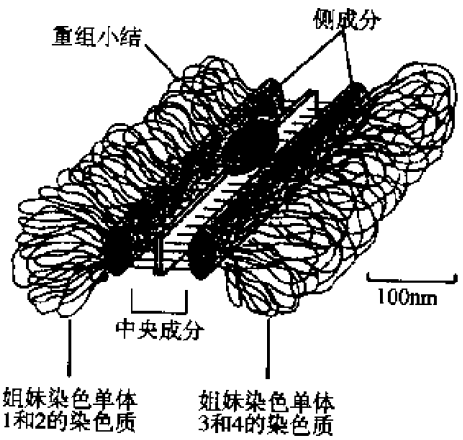


图1-18　联会复合体结构示意图

联会复合体已在各种动植物减数分裂前期观察到，它可能是真核生物减数分裂的共同特征。它起始于细线期，到双线期消失，在偶线期和粗线期最为清晰。电镜下它由三个平行的部分组成：两个平行的梯子样侧成分（lateral element），宽20nm～40nm，电子密度高。两侧成分间为一个含有中央成分（central element）的中央区（central space），宽100nm，比较明亮。中央成分宽300Å，比较暗。侧成分和中央成分之间有横向排列的纤

维相连称 LC 纤维(如图 1-19)。细胞化学分析表明,侧成分富含 DNA、RNA 和蛋白质(包括组蛋白),而中央成分主要含有 RNA 和蛋白质,没有或很少 DNA。虽然目前还没有直接证据证明 SC 的两条侧生成分是否参与交换的发生,但在 SC 的某些区段有时可观察到两条侧生组分之间有电子密度较高的结构,称为重组节(recombination nodule,RN)。联会复合体在双线期解体时,首先失去中央成分,然后侧成分消失。消失的原因可能是染色体进一步收缩的结果。

联会是染色单体之间遗传物质交换和减数分裂后期 I 同源染色体有规律分离的必要条件。

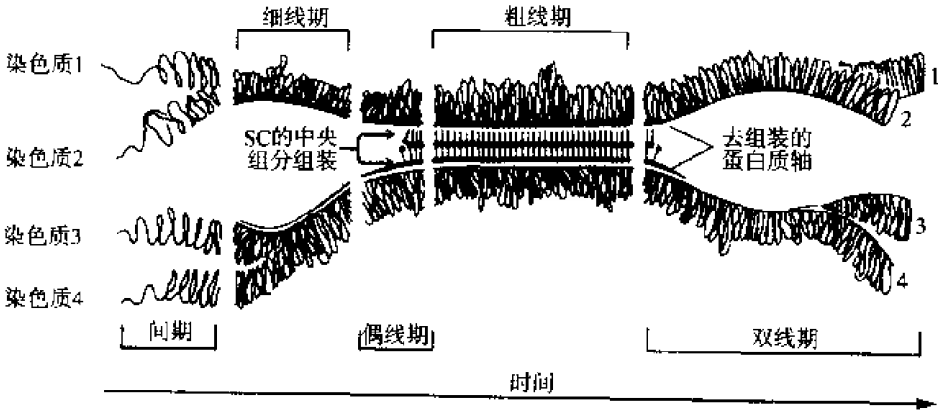


图 1-19 减数分裂前期 I 的进程与联会复合体

(3)粗线期(pachynema) 又称重组期(recombination stage)。该期始于同源染色体配对完成之后。染色体变粗变短,结合紧密,非姐妹染色单体之间发生 DNA 片段交换,即基因交换,其结果导致重组。人们对重组节作深入研究时发现,在 SC 的不同发育阶段处于 SC 中央组分上的重组节在形态、大小和分布特点方面并不完全相同,通常可以分为两种类型:第一种 RN 在偶线期 SC 片段刚刚形成时即出现,其数目在早粗线期达到最多,这一时期它们在不同染色体间和同一染色体上的分布是随机的,人们通常称这类重组节为早重组节(early RN);中-晚粗线期 SC 已稳定形成之后,第二种 RN 出现,其形态和大小与第一种 RN 不同,其分布也是不随机的,人们通常称这类重组节为晚重组节(late RN)。现在,人们还不敢明确断定,两种 RN 是各自独立形成的还是由早期重组节经形态分化形成晚期重组节。有一个支持分化过程的观察结果是,Bojko 在红色面包霉[粗糙脉胞菌(*Neurospora carassa*)]中发现了一种与早、晚重组节形态都不相同的 RN,Bojko 认为这可能是早重组节分化形成晚重组节过程中的一种中间过渡形式。现在人们已经有较多的证据支持晚重组节真正参与了交换,这些证据概括起来主要有:①像交换一样,RN 也出现在常染色质区;②交换的数目与 RN 的数目具有很好的一致性;③交换的位置与 RN 出现的位置具有很好的一致性;④影响交换数目的突变也能影响 RN 的数目,RN 频率的降低与交换频率的降低具有高度一致性;⑤像交换一样,RN 在分布上也是不随机的,存在干涉(interference)现象。

关于早重组节的功能,人们现在仍然知道得很少,但不同的研究者依据各自不同的观察结果提出了不同观点:①早 RN 参与了基因转换(gene conversion)。Holm 和 Rasmussen

曾提出，所有的早 RN 都具有参与交换和(或)基因转换的潜在可能，但只有那些最终参与交换的早 RN 能够始终保持与 SC 的联系，而参与基因转换的早 RN 则在偶线期-粗线期过渡时期脱离 SC，这样一来，早 RN 到晚 RN 形态上的变化也就反映了交换的发生。同样的观点，Carpenter 也曾提出过，果蝇中基因转换是由早 RN 作媒介完成的，而交换则由粗线期的晚 RN 作媒介来完成。②早 RN 参与了同源染色体的配对识别过程。Carpenter 曾提出"二步过程假说"(two-step process hypothesis)，认为早 RN 参与了为保证同源染色体精确配对而进行的同源性识别过程。具体来说，最初 SC 片段的形成并不要求同源性，而只是一个随机配对过程，这些片段形成后，由早 RN 参与立即通过一种类似基因转换的过程进行同源配对识别，当在刚形成的 SC 片段及其两侧能够识别出较长一段同源序列时，联会则继续进行。反之，当识别不同源序列时，SC 片段的两个侧成分则解除联系，然后再重复上述随机配对和同源识别过程。这个假说与传统观点不同之处在于，它认为同源识别发生在 SC 片段形成之后而不是之前。基因转换参与了同源染色体配对时同源识别过程的观点最初是由 Smithies 和 Powers 提出来的。他们在研究了位于不同染色体上的胎儿珠蛋白基因(foetal globin gene)间的基因转换过程后提出，在短同源序列间形成异源双链 DNA(同时产生基因转换)对于完成同源染色体识别来说似乎是一个很好的、可被自然选择下来的过程。有关重组的机制具体见遗传重组一章。

(4)双线期(diplonema)　在双线期，染色体进一步收缩变粗，同源染色体间有几个接触点，这时可清楚地看到四分体。同源染色体之间的接触点称为染色体交叉(chromosome chiasmata)，这也是从形态学角度提出的粗线期同源染色体之间发生交换的证据。在这一时期，染色体的联会引力结束，SC 消失，同源染色体在相斥中分离，而只在交换点连结在一起。随着双线期的进行，着丝粒分开且交叉点数目减少，这一过程称为交叉端化作用(terminalization)。当交叉移动到染色体末端时，就停留在那里，形成末端交叉，它似乎被阻在端粒内而不能越过这个结构，直到完成染色体中的中期取向排列，端部张力使末端交叉分开，而进入后期 I。

(5)终变期(diakinesis)　终变期是观察染色体的理想时期之一。唯一的缺点是这个时期比较短。由于染色体的高度螺旋化和二价体之间的相斥以及由于纺锤丝尚未与染色体结合，从而使染色体遍布整个细胞。

在终变期，核仁消失，四分体较均匀地分布在核中，姐妹染色单体靠着丝粒连在一起，而四分体只靠端部交叉结合在一起。由于交叉位置和数目的差异，端化程度不同，从而使终变期染色体呈现不同形状，如棒状、单环状，双环状或环叉状等。

2.中期 I

在第一次减数分裂中，核膜的消失，中心体的分离和纺锤体的形成，标志着前期 I 的结束。二价体的着丝粒与纺锤丝相连并向赤道板移动，每一对同源染色体的两个着丝粒向极取向，标志着减数分裂 I 已进入中期。

有丝分裂中期与减数分裂中期 I 有显著差异。在有丝分裂的中期，姐妹染色单体的着丝粒连在一起，并准确地位于赤道板上，同源染色体之间不发生配对。而在减数分裂中期 I 中，同源染色体的着丝粒并不位于赤道板上，而是位于赤道板两侧，由于配对而形成的末端交叉点位于赤道板上。当所有的二价体都到达这一位置时，则处于暂时的平衡状

态。二价体的每条染色体在赤道板两侧的向极分布是随机的，这种随机取向也决定了该染色体在子细胞中的分配，从而造成了染色体之间的不同组合方式（染色体重组）。如假设有 n 对同源染色体，则组合方式有 2^n 种。减数分裂分离是孟德尔自由组合定律的实质。所不同的是，孟德尔规律所研究的是基因分离，而减数分裂分离所涉及的是基因群的分离。

3. 后期 I

在后期 I，二价体中的两条同源染色体分开，分别向两极移动。减数分裂后期 I 与有丝分裂的差异是，在有丝分裂后期中，相连的姐妹染色单体的着丝粒分裂，每一染色单体作为独立的子染色体分别向两极移动。而在减数分裂后期 I 中，相连的姊妹染色单体的着丝粒并不分裂，两条染色单体作为一个单位向极移动。

4. 末期 I 及间期

同源染色体平均分配到达两极时，末期 I 开始。此时染色体的行为有两种类型：一是染色体脱螺旋而进入间期状态并形成核膜；另外一种是在纺锤体消失后，染色体并不解旋，而是由两极直接进入各自第二次分裂的赤道板。

在末期 I 后，随之而来的是一短暂的停顿时期，此时染色体不合成新的 DNA，染色体不存在再复制。

经过减数分裂 I 后，同源染色体平均地分配到子核中，从细胞的二倍体（$2n$）减少到子细胞中的单倍体（n），但其 DNA 含量与 G_1 期相同（2C），因为每条染色体含有两条姐妹染色单体。

（四）减数分裂 II

与有丝分裂过程基本相同，可分为前、中、后、末 4 期。在末期 I 和分裂间期染色体已经解旋，则前期 II 染色质重凝缩和螺旋化，每条染色体的两条染色单体在着丝粒处相连。在中期 II，染色体的着丝粒排列在赤道板上，每条染色体上的着丝粒分别和不同极的纺锤丝相连，每一着丝粒一分为二，每一条染色单体成为独立的子染色体，并在纺锤丝牵引下移向两极，其结果每极含有 1C 的 DNA 含量相当于减数分裂前间期 G_1 期细胞的 ½DNA 含量和 G_2 期细胞的 ¼DNA 含量。末期 II 重新组成核仁、核膜，染色体脱螺旋。经过减数分裂，使一个母细胞变成 4 个子细胞，染色体数目从母细胞的 $2n$ 变成子细胞的 n，从而完成减数分裂周期。

由减数分裂所形成的 4 个细胞核，不同性别其结果也不同。雄性动物经过减数分裂所形成的 4 个精细胞，分化成有功能的雄性配子——精子。而在雌性动物中，由于胞质的不均匀分裂，只形成一个卵细胞；其他子核形成胞质极少的无功能的极体（polar body）。从性染色体构成来说，精原细胞为 XY，卵原细胞为 XX，经过减数分裂后所形成的 4 个精子中性染色体分别为 X、X、Y、Y，而一个卵子的性染色体为 X。

（五）有丝分裂和减数分裂 DNA 含量和染色体数目的变化

染色体是 DNA 的载体，DNA 含量的加倍是染色体复制的一个象征。DNA 含量不仅在一个物种的某一发育阶段是恒定的，而且它与染色体的倍数水平密切相关。如果用 C 来代表单倍体中 DNA 的含量，那么一个二倍体细胞在合子或体细胞有丝分裂 G_1 期时，其 DNA 含量为 $2C$。由于在 S 期进行 DNA 复制，DNA 含量增加到 $4C$。当后期染色体单体分离后，在末期形成 2 个子细胞，细胞核内的 DNA 含量又恢复到 $2C$。处于连续有丝分裂的细胞中，其 DNA 含量总是表现出这种有规律的变化。在减数分裂 S 期中，DNA 含量也上升到 $4C$。由于染色体和 DNA 只复制一次，却进行了两次连续的分裂，所以在经过第一次减数分裂后期(I)后形成的子细胞中，核内 DNA 含量为 $2C$，第二次减数分裂后期(II)以后，核内 DNA 含量为 C。

染色体数目在有丝分裂和减数分裂中，也表现出有规律的变化。如果单倍体的染色体数目用 n 来表示，则合子和体细胞中染色体数目为 $2n$，经过 S 期后，虽然细胞核内其 DNA 含量加倍，由于着丝粒未分离，因此每条染色体含有两条染色单体，总的染色体数目却没有发生变化。在减数分裂过程中，真正的减数分裂是在后期 I 发生的，而有丝分裂后期及第二次减数分裂后期所形成的子细胞中染色体数目与该次分裂前细胞的染色体数目相同。DNA 含量和染色体数目的变化具体见图 1－20、图 1－21。

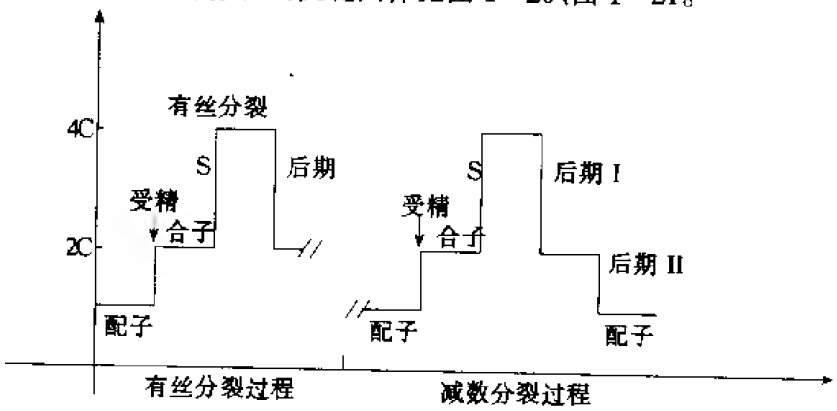


图 1－20　DNA 含量变化图解

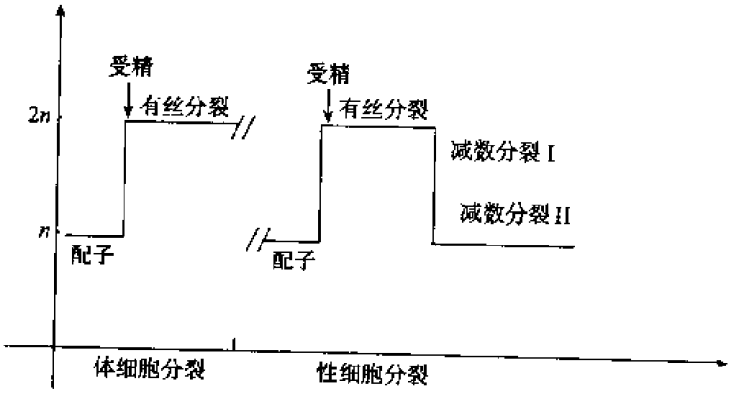


图 1－21　染色体数目变化图解

（六）细胞分裂的意义

1.有丝分裂

多细胞生物通过细胞分裂进行生长，因而有丝分裂也称体细胞分裂。在细胞分裂过程中，核内染色体经过复制并一分为二，均匀地分给子细胞，使子细胞获得与母细胞同样数量和质量的染色体。这种均等式分裂，可以使每个细胞都能得到与母细胞或当初受精卵同样的一套完整的遗传信息，这样既能维持个体的正常生长发育，也保证了物种的连续性和稳定性，因此具有重要的遗传学意义。

2.减数分裂

减数分裂对生物的遗传和变异起着十分重要的作用。

(1)维持物种染色体数目的恒定　　性母细胞($2n$)经减数分裂产生4个子细胞，以后分化成雌雄配子，各具有单倍染色体数(n)。当雌雄配子经受精作用结合成合子时，染色体数目又恢复成二倍($2n$)，从而保证了上代与下代之间遗传物质含量的恒定，为物种的延续提供了可靠的保证。

(2)为生物的变异提供了物质基础　　后期Ⅰ各对同源染色体的两个成员彼此分离，并向两极移动，但哪条同源染色体分向哪一极完全是随机的，这就有可能产生多种染色体组合的子细胞，如人体细胞为46条染色体，配子中含有的23条染色体每条都是一对同源染色体中的一条。即该配子的染色体构成只是2^{23}组合方式中的一种。因此这种染色体重组本身就为生物的变异提供了丰富的材料。

而且，同源染色体之间的交叉互换，为染色体上的基因互换提供了可能性，增加了子细胞在遗传组成上的多样性。因而减数分裂为生物的变异提供了物质基础，为自然选择和人工选择提供了丰富的材料，有利于生物的进化和新品种的选育。

第五节　染色体畸变[86~98]

自然界中各种生物的染色体都有一定的形态和数目，并通过精确的复制代代相传，从而保持染色体结构和数目的稳定性。但在自然条件下由于温度、营养、生理等异常变化，有可能使染色体的结构和数目发生变异。染色体畸变(chromosome aberration)是指在自然突变和人工诱变的情况下，染色体的结构和染色体数目发生的变化。由于染色体畸变的发生，使染色体上的基因数目和位置也发生了变化，从而导致个体性状表现型的决定性改变。

一、染色体结构变异

（一）结构变异的类型

染色体结构的改变取决于染色体的断裂和断裂后的愈合方式。如果断裂的染色体按

原有的顺序修补再恢复，这种现象称为染色体重建。如果断裂的染色体未发生重建，则产生染色体的结构改变。

染色体结构变异主要有 4 种类型：缺失(deletion)、重复(duplication)、倒位(inversion)和易位(translocation)，请参见图 1－22。

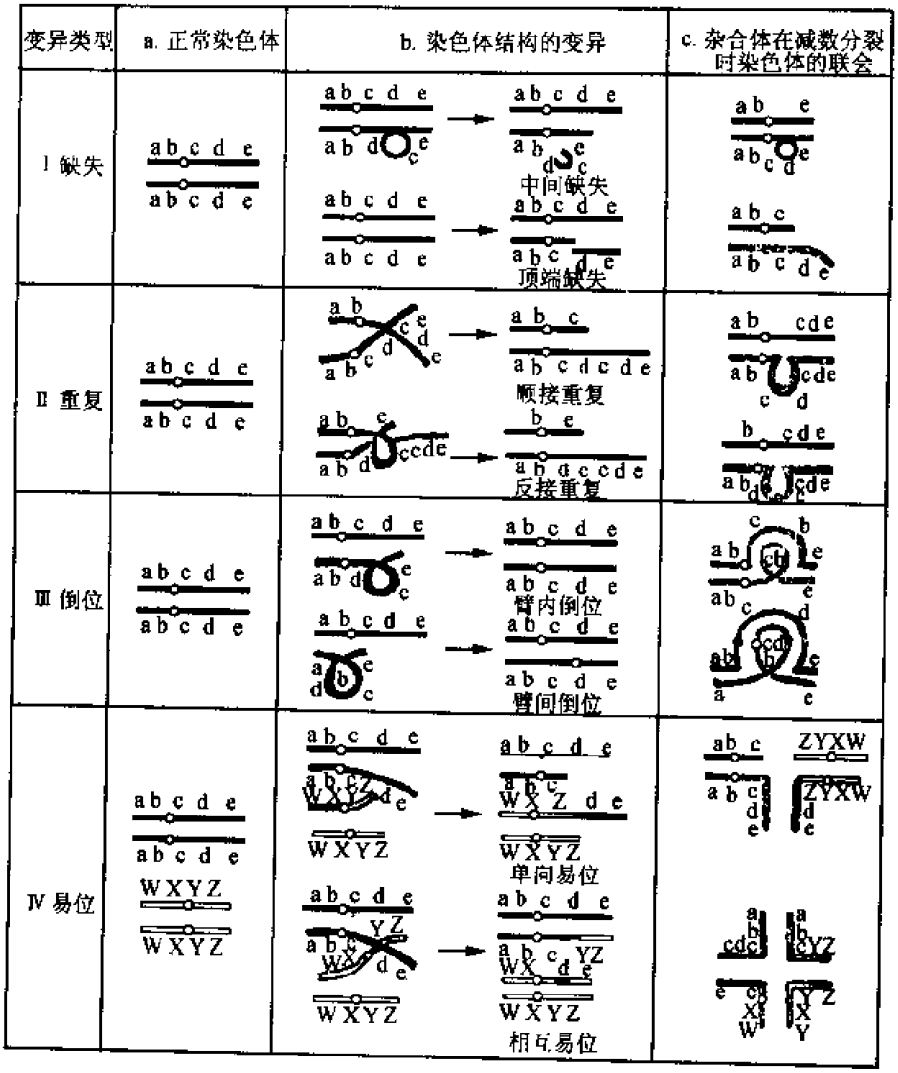


图 1－22　染色体结构变异类型及杂合子在减数分裂过程中的联会示意图

1. 缺失

缺失(deletion)是指一个正常染色体某区段的丢失，位于该区段上的基因也随之丢失。如果缺失发生在染色体的末端，称为末端缺失(terminal deletion)。末端缺失的发生较简单，只要在染色体的近末端发生一次断裂。如果缺失发生在染色体的内部，称为中间缺失(interstitial deletion)。中间缺失的发生较为复杂，要在染色体的某一臂上发生两次断裂并经过一次重接后，方可形成。由于末端缺失所形成的断头无端粒存在，因而很不稳定，其断头易与其他染色体断头相互连接，产生新的结构变异。因此染色体的中间缺失较

为常见。

发生了缺失的染色体称为缺失染色体。从染色体上断裂下来的无着丝粒片段称为断片。一对同源染色体只有一条发生缺失，这样的个体称为缺失杂合体。当减数分裂染色体联会时，由于缺失染色体不能与正常染色体相应地配对，因而在细胞学上可以观察到正常染色体上出现一个拱形的缺失圈（中间缺失）或多出一段（末端缺失）。

缺失对生物的发育是有影响的，影响程度取决于缺失片段的长短及其所载基因的重要性。一般来说，具有同一缺失的雌雄配子结合后的合子是很难发育的。自然界中一般不存在缺失纯合体。

缺失的遗传效应是破坏了正常的连锁群，影响基因间的交换和重组。而且当缺失了显性基因时，则杂合体表现出假显性现象（pseudo dominance）。在遗传试验中，利用缺失杂合体为材料，根据假显性现象和缺失圈部位，可以进行某些基因的定位。

2.重复

重复通常是指染色体上额外增加了与本身相同的某区段。如果重复区段上的基因排列顺序与原来相同，则称为顺接重复（tandem duplication）。如果重复区段上的基因顺序与原来相反，则称反接重复（reverse duplication）。

发生染色体重复的一种可能途径，是两条同源染色体均发生了断裂，且两条同源染色体中一条发生一次断裂，而另一条则发生两次断裂。在减数分裂的粗线期同源染色体发生了不等交换（unequal crossing over），可以产生重复，而且重复和缺失总是伴随出现的。重复染色体与正常染色体在减数分裂过程中进行联会时，可以在重复染色体上出现拱型的重复圈结构。重复圈与缺失圈在形态上难以区分，但在实质上，缺失圈出现在正常染色体上，而重复圈则出现在异常染色体上。

只要重复的区段不太大，重复对生物个体的生长发育并不一定带来有害的影响，这与缺失不同，缺失往往是有害的，会产生致死效应。重复的遗传效应同样是破坏了正常的连锁群，影响交换率，而且重复可以使基因的数量发生变化而引起剂量效应（dosage effect）。

3.倒位

一个正常染色体的某一区段断裂下来，后经180°的倒转后又重新接起来，成为一条具有倒位区段的染色体，这一现象称为倒位。倒位未增加任何遗传物质，只是改变了染色体上基因排列的顺序。倒位可以分为臂内倒位（paracentric inversion）和臂间倒位（pericentric inversion）两种。倒位发生在染色体臂内，倒位区段不包括着丝粒的，称为臂内倒位。如果倒位涉及两个染色体臂，倒位区段包括着丝粒的，称为臂间倒位。

臂内倒位和臂间倒位在减数分裂过程中，杂合体的两条同源染色体联会时，会出现倒位圈结构。而且当倒位圈内发生单交换时，可出现具有双着丝粒的染色体桥和断片（臂内倒位）。无着丝粒的断片往往留在胞质中而不能在细胞分裂过程中进入子细胞，而双着丝粒染色体在细胞分裂中被拉断并分别进入子细胞，但由于缺失区段大，配子往往死亡。最后存活下来的配子往往是没有发生过交换的配子。所以倒位可以抑制或大大降低杂合倒位圈内的基因重组。

4.易位

缺失、重复和倒位主要表现在同源染色体内结构的改变，而非同源染色体之间某区段的转移称为易位。根据染色体发生断裂的次数，可把易位分为4种类型：

(1)简单易位(simple translocation)　一条染色体的末端发生断裂并移接另一条非同源染色体的末端。这种现象极少见。因为正常染色体末端均具有端粒结构，而不与任何染色体或断头相连接，但缺乏端粒染色体则具有“黏性”。因此，简单易位可能也是相互易位的一种，即发生了两次染色体断裂。

(2)相互易位(reciprocal translocation)　即两个非同源染色体的每一条都有断点，然后相互交换断片。这是所观察到的易位的重要类型。

(3)移位易位(shift translocation)　断裂的染色体片段插入到非同源染色体或同一染色体的另一断裂位置上，称移位易位。

(4)复合易位(complex translocation)　复合易位是那些涉及三个以上断点的易位。Kaufmann曾用4000伦琴X射线处理雄果蝇并进行了144小时的红外线照射，发现其后代具有约32个断点的复合易位。

从细胞学上鉴定易位，也是根据易位杂合体减数分裂时染色体联会的特殊结构形式。减数分裂的粗线期，可以见到由4条染色体紧密配对形成的“+”字型结构。到了终变期，由于交叉移动而形成由4条染色体组成的四体环或四体链。

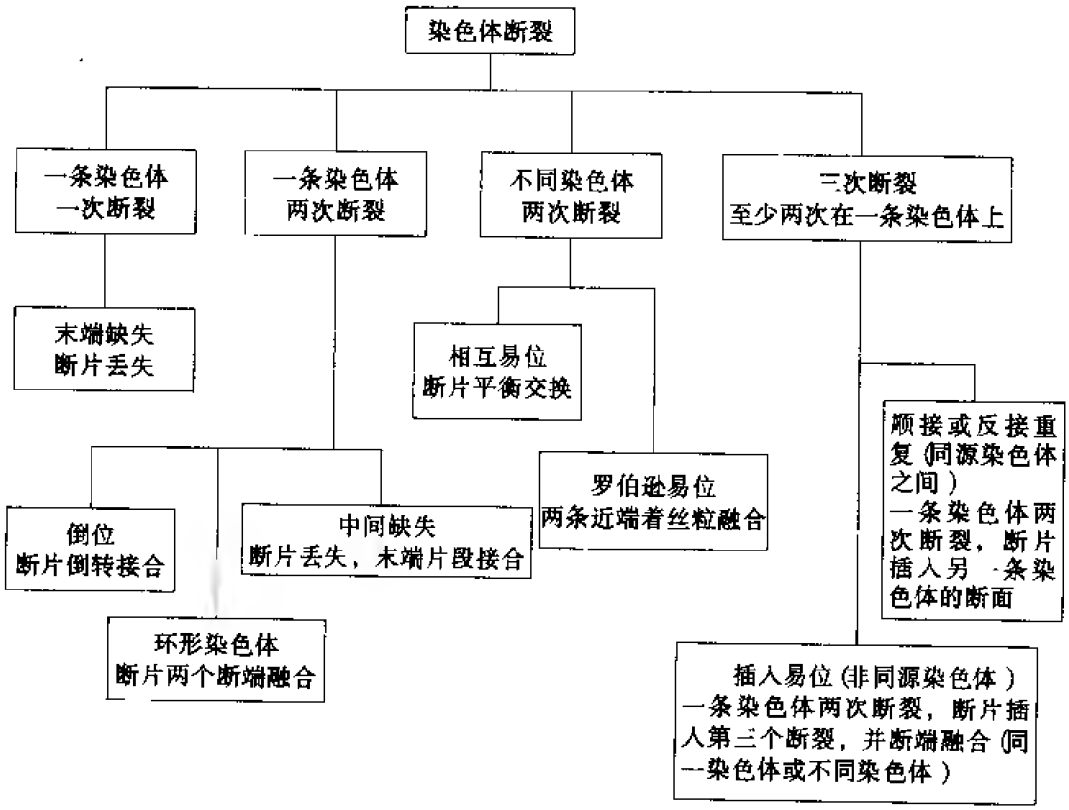


图1-23　染色体断裂与结构变异的关系

另外一种比较常见的易位方式是罗伯逊易位，即着丝粒融合型。由于两个端部或近端着丝粒融合，而使染色体数目减少。如西门答尔牛中发现的 1/21 染色体易位，家猪 13/17 染色体易位等。

（二）结构变异机制

染色体之所以发生结构变异，根本原因是在内因或外因的作用下，使染色体发生一次或一次以上的断裂。由于断裂后染色体末端端粒的丢失，断端具有与断片或其他断裂染色体断端重新接合的能力。当一个染色体断端与其本身的断片接合时，则不会发生结构变异。但如果一对同源染色体的不同区段发生断裂，而断裂区段的断片以不同方式进行接合，则会形成染色体缺失、重复或倒位等变异。当两对非同源染色体的各一条染色体断裂后，如它们的断裂区段间发生单向接合或相互接合，则会形成染色体易位。染色体断裂与变异的关系如图 1－23 所示。

除了上述变异之外，还有其他结构变异如等臂染色体（isochromosome）。它是一条异常的中着丝粒染色体，它的形成是由于近端着丝粒染色体复制后形成一对姐妹染色单体，在细胞分裂时着丝粒异常分裂（横向分裂），而使两条姐妹染色单体的两个长臂分离到一个子细胞而另外一子细胞只得到两条姐妹染色单体的短臂。

二、染色体的数目变异

染色体数目变异是指染色体数目发生不正常的变化。在讨论染色体数目变异时，首先应明确染色体组的概念。

（一）染色体组

二倍体生物配子中所包含的全部染色体称为染色体组。一个染色体组（n）包含一定数目的染色体，如马含有 32 条，人含有 23 条，猪有 19 条等。同一染色体组的各条染色体，其形态特征和基因连锁群都彼此不同，但它们构成了一个完整而协调的遗传体系，是生物生长发育所必需的遗传物质总体，如果缺少了其中任何一条，都会造成不育或性状变异，这是染色体组的最基本特征。

由于内外环境条件的影响，物种染色体组或其中成员的数目可能发生变化，变化的类型主要包括整倍体和非整倍体。

（二）整倍体变异

是指体细胞内含有完整的染色体组变异。按其染色组数变化，可分成以下几种：

1. 单倍体

单倍体（haploidy）是指体细胞内只含有一个染色体组。它具有正常体细胞染色体数

($2n$)的一半(n)。即它只含有单套的基因组。

在动物中有少数自然存在的单倍体。如膜翅目昆虫(蜂、蚁)和某些同翅目昆虫(白蚁)的雄性个体,是由未经受精的卵发育而成的单倍体。这些雄性个体虽然只含有单个染色体组,但正常可育,产生具有完整染色体组的精子。

2.多倍体

正常二倍体生物中体细胞内含有两个染色体组($2n$),含有三个染色体组的称三倍体($3n$)。凡是体细胞内含有三个以上染色体组的个体称为多倍体(polyploid)。根据染色体组的来源,多倍体又可分为同源多倍体(autopolyploid)和异源多倍体(allopolyploid)。来源相同并超过两个以上染色体组的个体称为同源多倍体。动物中的同源三倍体的形成主要是由于双精受精(dispermy)或异常的二倍体配子受精;四倍体的形成主要由于受精卵即合子(zygote)虽然进行了染色体的复制却没有进行分裂所致。在动物中组成型多倍体(constitutional polyploid)是很少见的,但几乎所有的动物个体中某些细胞总是存在着自然多倍体的现象。如人的巨噬细胞以及骨髓细胞等。另外其他的一些终末分化细胞(红细胞、鳞状上皮细胞)以及血小板没有细胞核,因此被称为缺倍体(nulliploid)。

来源于不同物种并超过两个染色体组的个体称为异源多倍体。如两个不同物种的二倍体生物杂交,它们的杂种再经过染色体加倍就可能形成异源四倍体或称双二倍体。如我国培育成的八倍体小黑麦就是人工创造的异源四倍体。

多倍体物种在植物中是很普遍的。因为大多数植物是雌雄同体或同花的。其雌雄配子常可能同时发生不正常的减数分裂,使配子中染色体数目不减半,因而通过自体受精而自然形成多倍体。许多植物可通过多倍体形成新物种,即多倍体是物种起源的方式之一。

动物界大多是雌雄异体,而雌 雄配子同时发生不正常的减数分裂机会很少,且染色体稍不平衡,就会导致不育,故动物界多倍体是很少见的。

(三)非整倍体变异

非整倍性(aneuploidy)变异是指在正常体细胞($2n$)的基础上发生个别染色体的增减现象。按其变异情况又可分为以下几种:

(1)亚倍体(hypoploid)　是指二倍体生物体中缺少一条或几条染色体。①单体性(monosomy),是指二倍体个体中缺少一条染色体($2n-1$);②缺对性(nullisomy)指在二倍体个体中缺少一对同源染色体($2n-2$)。

(2)超二倍体(hyperploid)　指比正常的二倍体染色体数多一条或几条染色体的个体。①三体性(trisomy),比正常的二倍体染色体数多一条染色体($2n+1$);②四体性(tetrasomy),比正常二倍体染色体数多一对同源染色体;③双三体性(ditrisomy),比正常的二倍体染色体数目多了两条染色体($2n+1+1$),与四体虽然染色体数目相同,但有本质的区别。

非整倍体的出现,主要有以下两个原因:

(1)染色体不分离(nondisjunction)　即在第一次减数分裂时,联会配对的两条染色体未分离而同时进入到子细胞中,或者在第二次减数分裂和有丝分裂过程中,姐妹染色单体不分离而同时进入子细胞中,而另一个子细胞则缺少了一条染色体。

(2)后期延迟(anaphase lag)　在正常情况下,细胞分裂后期的染色体移动是同步的。当一个染色体组的某个成员由于后期延迟运动(delayed movement,lagging)而未能随整个染色体组一起进入子细胞时,则会发生丢失现象,而使子细胞缺少这条染色体。

除整倍体变异和非整倍体变异外还有混倍性(mixoploidy)变异,即由两类不同染色体数目细胞所组成的个体。根据细胞来源又可分为镶嵌型(mosaicism),即细胞来源于同一合子,但由于变异而形成不同染色体数目的细胞系而形成的个体和嵌合型(chimerism),即细胞来源于不同合子而嵌合在一起形成的个体。

(赵志辉)

参 考 文 献

[1] Baldi M I et al Science, 1992. 255:1404~1408
[2] Cavalier-Smith T. Trends Genet, 1991. 7:145~148
[3] Condon C et al. Proc Natl Acad Sci USA, 1996. 93:6992-6997
[4] Dennis P. Cell, 1997. 89:1007~1010
[5] Curcio M J et al. Trends Genet, 1996. 2:436~438
[6] Attardi D et al. Cell, 1997. 89:859~866
[7] Garrett R A et al. Sys Appl Microbiol,1994.16:680~691
[8] Hacker J et al. Mol Microbiol, 1997. 23:1089~1097
[9] Olsen G J et al. Cell, 1997.89:991~994
[10] Johnsin P F et al. Nature,1983.302:681~687
[11] Palmer L R et al. J Bacteriol, 1994.176:3820~3823
[12] Phizicky E M et al. Trends Biochem Sci,1993.18:31~34
[13] Thompson L D et al. J Biol Chem, 1990.265:18104~18111
[14] Woese C R et al.Proc Natl Acad Sci USA, 1990.87:4576~4579
[15] Cavalier-Smith T. Nature, 1989. 339:100~101
[16] Gouy M et al. Nature, 1989. 339:145~147
[17] Joyce G F et al. Nature, 1989. 338:217-224
[18] Pool R. Science, 1990. 247:159
[19] Woese C R et al. Sci Amer, 1981. 244:98~125
[20] Belfort M et al. Cell, 1987. 89:1003-1009
[21] Lee C et al. Ann Rev Cell Biol, 1986. 2:315~336
[22] Wickner W T et al. Science, 1986. 230:400~407
[23] Ellis R J et al. Ann Rev Biochem, 1991. 60:321~347
[24] Baker K P et al. Nature. 1991. 349:205~208
[25] Forbes D J. Ann Rev Cell Biol, 1992. 8:492~527
[26] Jentsch S. Ann Rev Genet, 1992. 26:179~207
[27] Georgopoulos C et al. Ann Rev Cell Biol, 1993. 9:601-634
[28] Ciechanover A. Cell, 1994. 79:13~21
[29] Walter P et al. Ann Rev Cell Biol, 1994. 10:87~119
[30] Davis L I. Ann Rev Biochem 1995. 64:865~896
[31] Gorlich D et al. Science, 1996. 271:1513~1518
[32] Schatz G. Science, 1996. 271:1519~1526
[33] Rapoport T A et al. Ann Rev Biochem, 1996. 65:271~303
[34] Coux O et al. Ann Rev Biochem, 1996. 65:801~848
[35] Stillman B. Cell, 1986. 45:555~565
[36] Burke D T et al. Science, 1987. 236:806~812
[37] Staudt L M et al. Science, 1988. 241:577~580

[38] Schleif R. Science, 1988. 241:1182~1187
[39] Drlica K et al. Microbiol Rev, 1987. 51:301~319
[40] Brock T D. Microbiol Rev, 1988. 52:397~411
[41] Black L W. Ann Rev Microbiol, 1989. 43:267~292
[42] Zakian V A. Ann Rev Genet, 1989. 23:579~601
[43] Schulman I et al. Ann Rev Cell Biol, 1991. 7:311~336
[44] Blackburn E H. Nature, 1991. 350:569~573
[45] Blackburn E H. Ann Rev Biochem, 1992. 61:113~129
[46] Hyman A A et al. Ann Rev Cell Biol, 1995. 11:471~495
[47] Zakian V A. Science, 1995. 270:1601~1607
[48] Greider C W. Ann Rev Biochem, 1996. 65:337~366
[49] Williamson J R et al. Cell, 1989. 59:871~880
[50] Shippen-Lentz D et al. Science, 1990. 247:546~552
[51] Craig J M et al. Bioessays, 1993. 15:349~354
[52] Dobbs D L et al. Nucl Acid Res, 1994. 22: 2479~2489
[53] Saitoh Y et al. Cell, 1994. 76:609~622
[54] Tyler-Smith C et al. Cuur Opin Genet Dev, 1993. 3:390~397
[55] Craig J M et al. Nature Genetics, 1994. 7:375~381
[56] Scott J. Cell, 1995. 81:833~836
[57] Zhang Q et al. J Hered, 1998. 89:348~353
[58] Larramendy M L et al. Hereditas, 1993. 119:89~94
[59] Hartwell L H et al. Science, 1989. 246:622~628
[60] Mcintosh J R et al. Science, 1989. 246:622~628
[61] Murray A W et al. Science, 1989. 246:614~621
[62] Nurse P. Nature, 1989. 344:503~508
[63] Belmont L D et al. Cell, 1990. 62:579~589
[64] Forsberg S L et al. Ann Rev Cell Biol, 1991. 7:227~256
[65] Ellis R E et al. Ann Rev Cell Biol, 1991. 7:663~698
[66] Norbury C et al. Ann Rev Biochem, 1992. 61:441~470
[67] Reed S J. Ann Rev Cell Biol, 1992. 8:529~561
[68] Herr C J. Cell, 1993. 73:1059~1065
[69] Nurse P. Cell, 1994. 79:547 -550
[70] Sherr C J. Cell, 1994. 79:551~555
[71] King R W et al. Cell, 1994. 79:563~571
[72] Sherr C J et al. Genes & Dev, 1995. 1149~1163
[73] Weinberg R A. Cell, 1995. 81:323~330
[74] Bloom K. Cell, 1993. 73:621~624
[75] Sawin K E et al. Bioessays, 1993. 15:399~407
[76] Holloway S. Cell, 1993. 73:1393~1402
[77] Surana U et al. EMBO J, 1993. 12:1969~1978
[78] Li R et al. Cell, 1991. 66:519~532
[79] Nigg E A. Cell Biol, 1992. 3:245~253
[80] McConnell S J et al. Science, 1993. 260:687~689
[81] Putnam A J et al. J Cell Sci, 1998. 111:3379~3387
[82] Dowd S et al. J Cell Sci, 1998. 111:3389~3399
[83] Green D R et al. Cell, 1998. 94:695~698
[84] Bruce A et al. Molecular Biology of The Cell. Garland Publishing Inc, 1989.
[85] Strachan T et al. Human Molecular Genetics. BIOS Scientific Publishers, 1989.
[86] Lewin B. Genes. New York:Oxford University Press, 1997
[87] 解生勇. 细胞遗传学. 北京:北京农业大学出版社,1989
[88] 刘祖洞等. 遗传学. 北京:高等教育出版社, 1991

[89] 吴仲贤等. 动物遗传学. 北京:农业出版社, 1980
[90] Bauchan G R et al. J Hered, 1994. 85:18~22
[91] Ducos A et al. J Hered, 1998. 89:136~142
[92] Astachova N M et al. Genet Sel Evol, 1991. 23(suppl):65~69
[93] Benkhalifa M et al. Ann Genet, 1992. 35:61~64
[94] Ducos A. Rev Med Vet, 1995. 146:715~722
[95] Ducos A et al. Rev Med Vet, 1995 147:101~108
[96] Yang H et al. Cytogenet Cell Genet, 1992. 61:67~74
[97] Kuokkanen M T et al. Hereditas, 1988. 106:147~149
[98] Gustavsson I et al. Hereditas, 1992. 117:31~38

第二章　遗传的物质基础

第一节　遗传物质的本质[1~4]

一、DNA 是遗传物质

核酸(nucleic acid)是重要的生物大分子,其研究是分子生物学的重要领域。核酸的研究已有一百多年历史,早在 1868 年,瑞士的一位科学家 F. Miescher 从外科绷带上的脓细胞核中分离出一种有机物质,其含磷量之高超过任何当时已经发现的有机化合物,并且有很强的酸性。由于这种物质是从细胞核中分离出来的,当时就称它为核素(nuclein)。Miescher所分离到的核素就是我们今天所指的脱氧核糖核蛋白。以后陆续证明,任何有机体包括病毒、细菌、动植物等,都无例外地含有核酸。核酸是一种线形多聚核苷酸,其基本结构单位是核苷酸,通过 3′,5′磷酸二酯键连成核酸链。核苷酸由碱基、戊糖、磷酸构成。在核酸中发现两种戊糖,根据含戊糖种类的不同把核酸分为两种,DNA 中的戊糖是 2-脱氧核糖,而在 RNA 中是核糖。核酸的主要生物学作用是遗传信息的载体。

因为遗传物质是通过表达多种多样的蛋白质实现其功能的,而早期人们错误地认为遗传物质的结构必然与其表达的蛋白质的结构同样复杂,所以在很长一段时间内人们认为,只有蛋白质才具有足够的多样性来确定其他的蛋白质,直到人们意识到遗传物质携带的是以密码形式存在并决定蛋白质组成的遗传信息时,才抛弃了这个错误的看法。

DNA 是遗传物质的概念起源于 1928 年 Griffith 等进行的转化(transformation)实验。在实验中,光滑型(S)的肺炎球菌对小鼠有致死性,而粗糙型(R)的肺炎球菌无致死性。单独存在的 R 与加热致死的 S 肺炎球菌均不能杀死小鼠,但当二者混合在一起却是致死的。利用人 R 和 S 抽提液的混合物也能观察到致死的结果,表明在抽提液中含有一种转化因子(transforming factor)。1944 年 Avery 等从 S 抽提液中部分纯化了转化因子,证明它是 DNA。利用此 DNA 样品加入 R 型细菌的培养物中,得到的菌落中含有了 S 型球菌。这个实验证明了使肺炎球菌的遗传性发生改变的转化因子是 DNA,而不是蛋白质。这一发现极大地推动了对核酸结构和功能的研究。

1952 年 Hershey 和 Chase 等利用大肠杆菌噬菌体的捣碎实验确证了 DNA 的遗传物质的本质。在噬菌体中,DNA 是唯一含磷的物质,而蛋白质是唯一含硫的物质。利用含 P 或 S 的放射性培养基培养噬菌体,得到了放射性标记的噬菌体。使标记的噬菌体吸附细菌,几分钟后离心除去未吸附的噬菌体。然后利用捣碎机捣碎使噬菌体和细菌分开。离心后细菌在沉淀中,而噬菌体在上清液中。这时发现放射性的硫有 80% 在上清液中,只有 20% 在沉淀中,而磷的情况则相反。说明在噬菌体感染的过程中,DNA 进入了细菌体内,而蛋白质留在细菌外。证明具有遗传作用的是 DNA 而不是蛋白质(图 2-1)。

DNA 是由 4 种基本的核苷酸组成(A、G、C、T)。1950 年以前,流行四核苷酸结构学说。该学说认为任何核酸分子都是由等摩尔的 4 种核苷酸组成,因此核酸不大可能具有

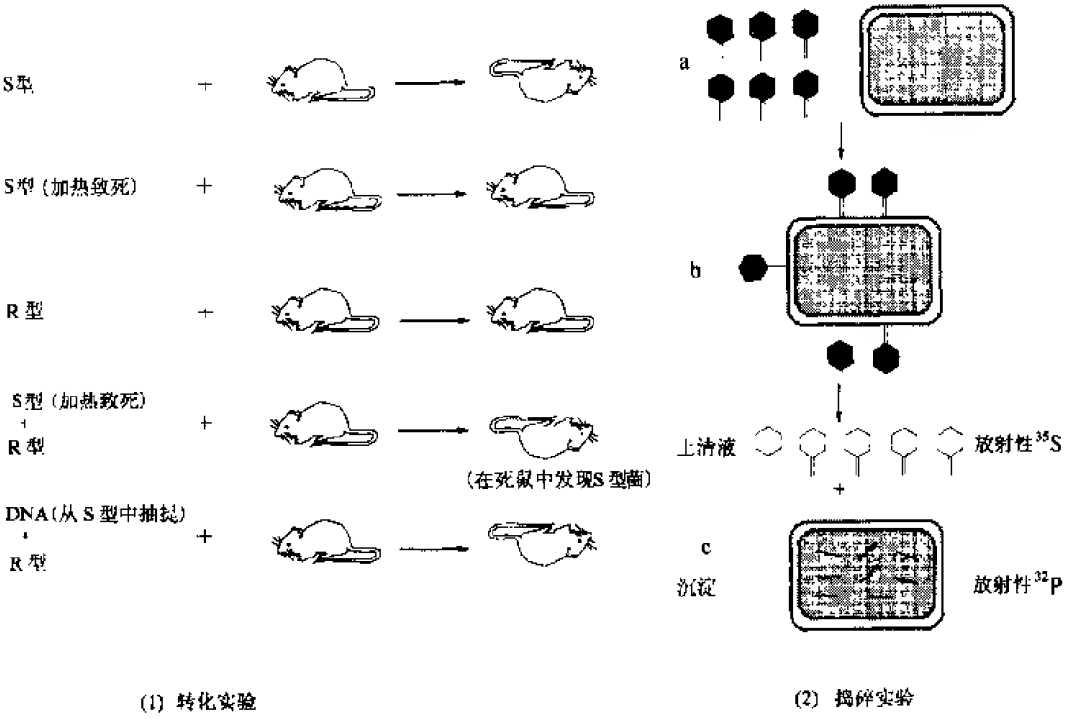


图 2-1 证明 DNA 是遗传物质的实验

(1)转化实验；

(2)捣碎实验。

a. 用标记^{32}P(标记 DNA)或^{35}S(标记蛋白质)的噬菌体感染细菌；b. 标记的噬菌体吸附细菌；c. 捣碎并离心使噬菌体蛋白质外壳与被感染的细菌分开。^{35}S 在上清液中，^{32}P 在沉淀中。

重要的生物功能。尽管 Avery 等的实验已证明了 DNA 是重要的遗传物质，反对者仍然认为只有蛋白质才是转化因子。1950 年以后，Chargaff、Markham 等应用纸层析及分光光度法大量测定了各种生物的 DNA 碱基组成后，发现不同生物的 DNA 碱基组成不同，有严格的种的特异性，这给四核苷酸学说以致命的打击。同时他们还发现，尽管不同生物的碱基组成不同，但其量总是 A=T、G=C，提示了 A-T、G-C 之间互补的概念。这一重要的发现，为以后 Watson 和 Crick 建立 DNA 双螺旋结构模型提供了重要依据。

1953 年 DNA 双螺旋结构模型的提出，被认为是本世纪自然科学的重大突破之一。它揭开了分子生物学研究的序幕，为分子遗传学的研究奠定了基础。

遗传物质必须具有以下特性：①贮存并表达遗传信息；②能把遗传信息传递给子代；③物理和化学性质稳定；④有遗传变化的能力。

DNA 具有上述特性，适合作为遗传物质。DNA 序列由不同的核苷酸排列而成，遗传信息就存在于其中。生物体 DNA 上的遗传信息通过表达产生许多特殊蛋白质来实现其功能。蛋白质由一系列特殊的氨基酸组成，每种蛋白质的结构和活性由其氨基酸排列的顺序决定，DNA 的核苷酸序列能编码组成它所对应蛋白质的氨基酸序列。DNA 序列和相应的蛋白质的氨基酸序列之间的关系叫做遗传密码(genetic code)。通过确定每种蛋白质中的氨基酸序列，一段 DNA 可以携带确定一个有活性的多肽链的所有信息。通过这

种方法，DNA 这个简单的结构就能够以无数的蛋白质的形式来表现它自己。综合不同蛋白质产物的结构和活性，生物体的细胞就可以表现出独特的性质。

DNA 在履行其遗传物质的使命时既具有稳定性又具有灵活性。DNA 是通过碱基互补配对形成的双链分子，碱基互补是其复制、转录、表达遗传信息的基础。复制过程中，通过碱基互补配对机制把遗传信息准确地传递给子代，而且 DNA 在生理状态下很稳定，这一切都表明 DNA 适于作为遗传物质。DNA 上也偶然会发生突变(mutation)，这些突变往往可稳定地遗传，这是生物进化的基础。

二、RNA 也可以作为遗传物质

上述特征对原核生物和真核生物细胞携带的遗传信息甚至病毒携带的遗传信息的表达都是同样适用的。所有已知的有机体和许多病毒的遗传物质都是 DNA 双螺旋。然而，一些病毒也采用另一种核酸- RNA 作为遗传物质。尽管 RNA 的化学结构式与 DNA 有所不同，它仍可发挥与 DNA 相同的功能。不同的病毒体系复制其核酸的机制的细节是不同的，但复制的基本原理是一样的，都通过其互补链的合成来实现。病毒的传染媒介是病毒颗粒(virion)，由病毒基因组及包被在外的蛋白质外壳组成。病毒的核酸编码包装基因组所需的蛋白质，还编码一些在感染循环中复制病毒所需的蛋白质，宿主细胞不能提供这些蛋白质；而病毒所需的其他蛋白质需要宿主细胞提供，因此病毒不能独立生存。

类病毒(viroid)是使高等植物产生疾病的有传染性的因子(agent)，它们是很小的环状 RNA 分子，与病毒不同，类病毒的 RNA 本身就是一个感染因子，类病毒只由 RNA 组成，其中广泛存在不完全的碱基配对，形成一种特有的棒状结构。能破坏这个棒状结构的突变可降低感染能力。类病毒有很多种。

类病毒与病毒类似，都有可遗传的核酸基因组，类病毒 RNA 的序列可忠实地传递给其后代。其核酸可以突变，并决定其表型，因此符合遗传物质的标准。但类病毒在结构和功能两方面都与病毒不同，它们有时被称为亚病毒病原体(subviral pathogens)。

类病毒的 RNA 似乎并不能翻译成蛋白质，即它不能亲自编码生存所需要的蛋白质，这导致了两个问题，类病毒的 RNA 是如何复制的？它是如何影响被感染的植物细胞的表型的？类病毒的复制必须由宿主细胞的酶来完成，其 RNA 可以作为复制模板。类病毒通过复制占有宿主细胞中关键的酶，从而影响细胞正常功能，所以它们通常是致病的。类病毒也能影响必需的细胞 RNA 的产生而引发疾病，它们还可以作为一个不正常的调控分子，对个别基因的表达产生特殊的影响。

三、是否存在核酸之外的其他遗传物质

综上所述，大部分生物的遗传物质是 DNA，而 RNA 病毒及类病毒的遗传物质是 RNA。但自然界中似乎有特例表明，还有除核酸之外的其他遗传物质。除了传统意义上的遗传物质外，一个更独特的"遗传因子"是羊搔痒病(scrapie)的致病因子，它是导致山羊和绵羊的退行性神经疾病的原因。这种病与影响人脑功能的库鲁病(kuru，新几内亚震颤病)和克-雅氏病(Creutzfeldt-Jakob disease，CJD)及疯牛病类似。这种因子具有可感染

性，例如把库鲁病人的脑组织均浆接种至黑猩猩脑内，可导致其发病。与病毒相似，搔痒病的感染性粒子是滤过性的，但与病毒不同的是，福尔马林处理不能完全消除其传染性。它可被蛋白酶 K 灭活，但不能被核酸酶处理和 UV 辐射灭活，表现出典型的蛋白质性质。在高度纯化的感染性物质中检测不到大于 100nt 的核酸，所以这种感染性粒子不是核酸。因为目前唯一鉴定的组分是蛋白质，所以这个异常的因子被命名为朊病毒(prion)，意为蛋白质样的感染性粒子。

它是一个 28 kDa 的疏水性的糖蛋白，称为朊病毒蛋白(prion protein，PrP)。PrP 由宿主细胞核基因(在哺乳动物中是保守的)编码，在正常动物的脑中也有表达。这个蛋白质以两种形式存在，在感染性和 PrP 蛋白的不同形式之间有紧密的联系。正常的脑组织中发现的产物是 PrP^c，它可以被蛋白酶完全降解。而另一种 PrP：PrP^{sc}只存在于被感染的脑组织中，它对蛋白酶的降解有较强的抗性，只能被部分降解。PrP^c 是一个可溶性的蛋白质，而 PrP^{sc}很难被溶解。两种蛋白质具有相同的一级结构，但其二级结构却十分不同，PrP^c 的 α 螺旋的含量约为 40%，几乎没有 β 片层结构，相反 PrP^{sc}含有高达 50% 的 β 片层，只有 20% 的 α 螺旋。结构上的差异可能导致了对蛋白酶抗性的不同。

搔痒病的发生伴随着细胞中 PrP^c 到 PrP^{sc}的转化现象。是什么导致了朊病毒疾病呢，是 PrP^c 的缺失还是 PrP^{sc}的产生？目前不能确切知道 PrP^c 是否有生物学功能，实验表明 PrP^c 的缺失会导致神经细胞的死亡，但不会导致经典的朊病毒疾病。那么是否 PrP^{sc}会导致疾病呢？实验表明 PrP^{sc}本身不足以产生疾病，PrP^{sc}只有产生于细胞内或者进入一个已经表达 PrP^c 的细胞时才有毒性，然后才能导致疾病的发生。缺少 *PrP* 基因的鼠不能被感染并发病，证明了 PrP^c 对疾病的发生是必不可少的，所以推测搔痒病的发生是感染性的 PrP^{sc}与内源的 PrP^c 共同作用的结果。

通过对两个鼠细胞株的观察发现内源的 PrP^c 可能参与了感染的发生：这两个细胞株 PrP^c 编码基因的两个位点存在差异，而朊病毒对它们的感染时间的长短是不同的。朊病毒感染的宿主专一性和种间障碍的现象表明 PrP^c 和 PrP^{sc}的相互作用可能介导了感染的转化。种间障碍是指一种生物对另一种生物产生的朊病毒的感染有抗性。通常小鼠对仓鼠的朊病毒有抗性，但通过转基因方法引入了仓鼠的 *PrP* 基因后，小鼠就对仓鼠的朊病毒变得敏感。如果小鼠的内源 *PrP* 基因失活会导致小鼠对自身的朊病毒产生抗性。表明在感染性的 PrP^{sc}和内源的 PrP^c 之间存在直接的相互作用，并且这种相互作用取决于 PrP 序列间的差异。所以传染性粒子产生于内源的 PrP^c 蛋白，内源的 PrP^c 可能是传染性的 PrP^{sc}的靶子，通过有感染性的 PrP 分子(PrP^{sc})的活化，使之由无害的结构转化为可感染的 PrP^{sc}结构。但纯化的 PrP^{sc}活性很低，并且在有些感染病例中未发现有蛋白酶抗性的物质，所以感染性 PrP 的结构也可能不同于蛋白酶抗性的 PrP^{sc}结构。

为了证明朊病毒本身就具有遗传物质的特征，还需证明由 PrP^c 到 PrP^{sc}转变在体外可以发生并能产生具有感染性的产物。虽然目前还不能证明这点，然而实验证明了在体外 PrP^{sc}可以把 PrP^c 转化为一种类似 PrP^{sc}的对蛋白酶水解有抗性的形式，同时通过体外反应也可重现体内转化反应的种属专一性，甚至细胞株的专一性。

人们对 PrP^c 转化为 PrP^{sc}的机制提出了不同的模式(图 2-2)，其共同点都认为：PrP 蛋白的错误折叠形式可以催化一个天然 PrP 分子转化为完全不同的折叠形式。在成核聚合 (nucleation-polymerization)模式中，PrP^{sc}的生成是一个依赖于晶核生成的聚合过程。

在事先没有聚集体存在时，PrP^c 与 PrP^{sc} 之间的转化是可逆的，但 PrP^{sc} 没有 PrP^c 稳定，PrP^{sc} 聚集体可与 PrP^{sc} 单体结合并使之稳定。感染缓慢的发生，如同结晶过程中引入晶种并生成晶体一样。在模板辅助(template-assisted)模式中，PrP^{sc} 形式比 PrP^c 稳定，但由 PrP^c 到 PrP^{sc} 的转化在动力学上是不利的。PrP^{sc} 可以催化 PrP^c 分子的结构的变化。它可结合并稳定 PrP^c 到 PrP^{sc} 的中间体，通过增加不稳定的中间体的数量或提高中间体自发地转化为 PrP^{sc} 的速率，促进生成更多稳定的 PrP^{sc}。在生物体中这两个模型都能找到先例，成核聚合模型类似于微管蛋白的聚集、晶体的生成、镰刀型血红蛋白的形成、病毒衣壳粒的组装、细菌鞭毛的聚集等生物过程。模板辅助模型也有先例，一些蛋白质都有动力学控制的构型转化，如流感血球凝集素、枯草杆菌蛋白酶等。

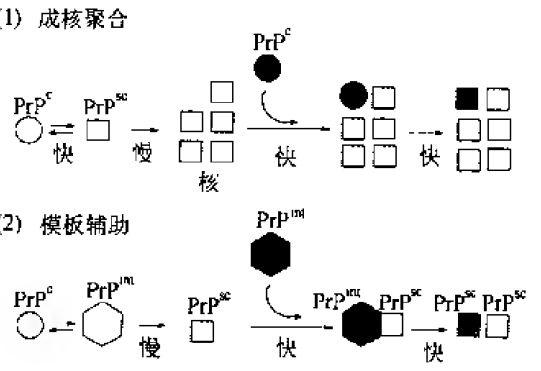


图 2-2 PrP^c 转化为 PrP^{sc} 的机制
(1)成核聚合机制； (2)模板辅助机制。

大量的数据表明了许多可转移性的神经疾病中 PrP 蛋白的中心地位。这些数据表明疾病的发生需要 PrP 蛋白，而且 PrP 蛋白从正常的可溶性的 α 螺旋构象向不溶性的 β 片层构象的转化最终导致了疾病和感染。但大多数转化的过程还不清楚，尤其是 PrP^c 存在时引入 PrP^{sc} 是否就可以导致感染？构象变化的本质是什么？这些问题都有待于进一步研究。

朊病毒假说还远远没有证实。目前还不能准确地鉴定朊病毒的性质，PrP 蛋白的正常功能和致病机理尚不清楚。朊病毒假说基于 PrP 结构变化的性质，认为不正常的形式(PrP^{sc})是可感染的。但是尽管感染中都有 PrP^{sc} 存在，但并不是所有的 PrP^{sc} 都导致感染的发生。

朊病毒假说面临的最大困难是不能满意地解释这类疾病的多菌株现象。不同来源的感染性物质，携带特定的遗传信息，可以产生不同的发病特征，如不同的感染时间，PrP^{sc} 蛋白酶抗性的不同。同一个宿主可被多种感染性物质感染，两种不同的 PrP^{sc} 株可以指导一种 PrP 结构转化为两个不同的构型，尽管宿主 PrP 蛋白一级结构没有差异。而同一个感染性物质，如牛的海绵状脑病粒子，可以在具有不同 PrP 蛋白的多种生物中传代(包括人)，即使经过多次传代仍然保持自身的生物学特性。目前没有一种机制可以满意地解释一种 PrP 蛋白怎样反映并保持多种感染性物质的生物学特性。人们假设存在一种不依赖于宿主的小的信息分子，可以结合宿主的 PrP 蛋白质并被其保护，能够携带菌株专一性的信息。人们推测它可能是一个不能翻译的核酸分子。如果发现了这种菌株专一性的信息分子，就可以推翻 PrP 蛋白唯一的假说。即使朊病毒假说最终被证实，是否足以证明朊病毒是一种遗传物质呢？PrP^{sc} 蛋白诱导的 PrP^c 的构型变化能被看成是遗传物质复制的一种形式吗？这些问题都有待于进一步的讨论。

Q75
ZYJ

第二节　核酸的化学组成

核酸是以核苷酸为基本结构单位，通过3′,5′磷酸二酯键(phosphodiester linkage)形成的链状多聚体。核苷酸(nucleotide)由含氮碱基(nitrogenous base)、戊糖、磷酸构成(图2-3)。

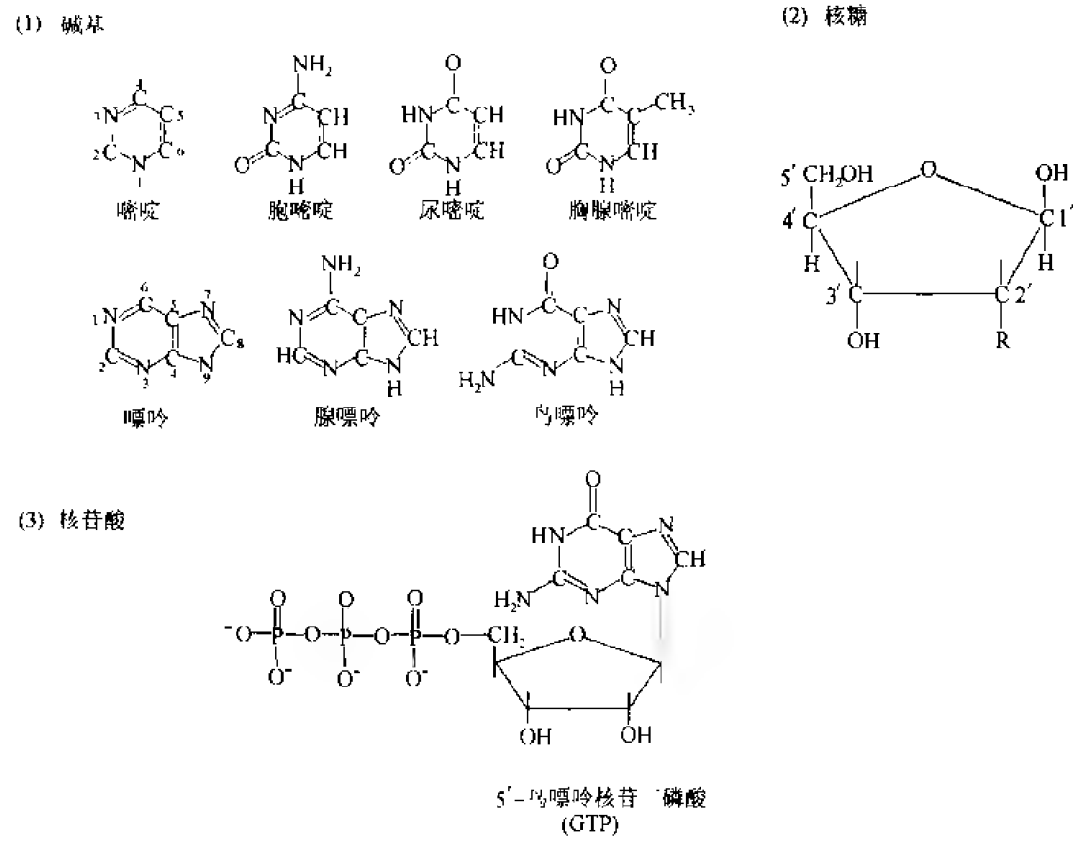


图2-3　碱基、核糖、核苷酸

(1)碱基；(2)核糖 R=OH 为核糖 R=H 为脱氧核糖；(3)5′-核苷三磷酸。

1.含氮碱基

核酸中的碱基分为两类：嘧啶碱和嘌呤碱。常见的嘧啶(pyrimidine)有三类：胞嘧啶(cytosine,C)、尿嘧啶(uracil,U)、胸腺嘧啶(thymine,T)；常见的嘌呤碱(purine)有两类：腺嘌呤(adenine,A)和鸟嘌呤(guanine,G)。除了5种基本的碱基外，核酸中还有一些含量甚少的碱基，称为稀有碱基。DNA和RNA各有4种碱基，参与DNA组成的4种碱基是腺嘌呤、鸟嘌呤、胞嘧啶、胸腺嘧啶；参与RNA组成的4种碱基是腺嘌呤、鸟嘌呤、胞嘧啶、尿嘧啶。Chargaff等在50年代对各种生物DNA的碱基组成进行了定量测定，发现所有DNA中腺嘌呤与胸腺嘧啶的摩尔含量相等(A=T)；鸟嘌呤与胞嘧啶的摩尔含量相等(G=C)。因此，嘌呤的总含量与嘧啶的总含量相等，即A+G=C+T。这一规律提示A与T,G与C之间碱基互补的可能性，为DNA双螺旋结构模型的建立提供了重要依据，称为Chargaff法则(Chargaff's rule)。

2.核苷

核苷(nucleoside)由戊糖和碱基缩合而成,并以糖苷键(glucoside bond)相连接。RNA链中的戊糖是D-核糖,DNA链中的戊糖是D-脱氧核糖。因此核苷可分为核糖核苷、脱氧核糖核苷。核糖核苷有4种:腺苷、鸟苷、胞苷、尿苷。脱氧核糖核苷也是4种:脱氧腺苷、脱氧鸟苷、脱氧胞苷、脱氧胸苷。

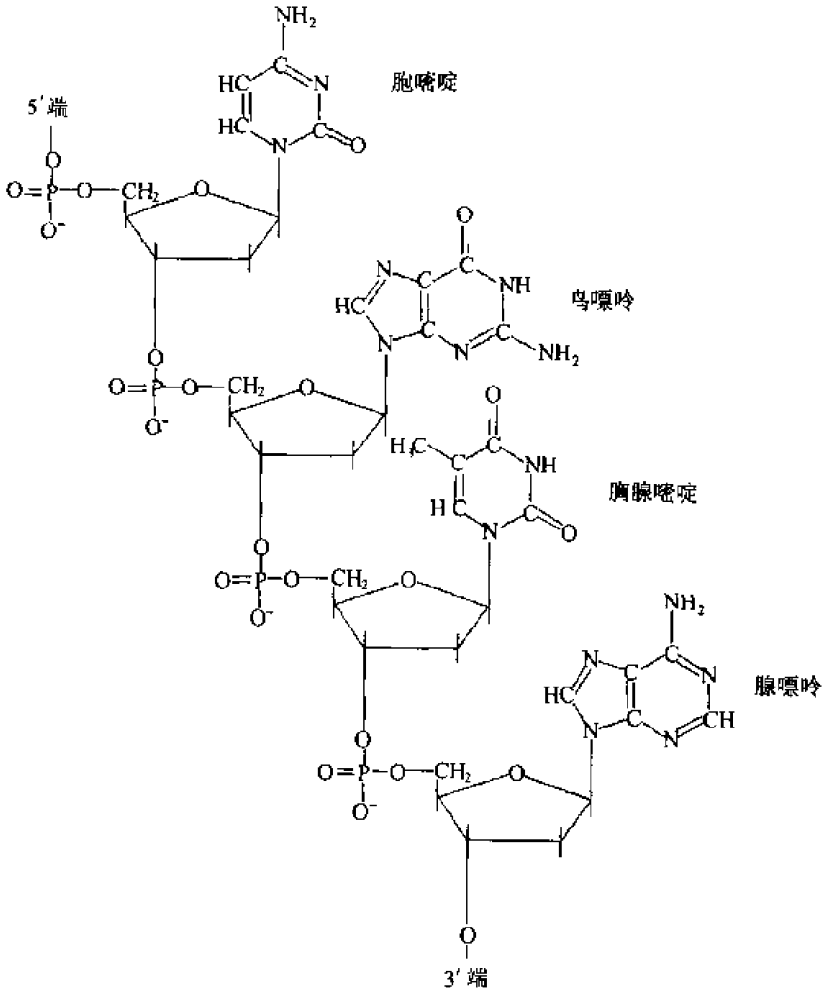


图2-4 DNA分子的一级结构

3.核苷酸

核苷的磷酸酯叫做核苷酸,分成核糖核苷酸和脱氧核糖核苷酸两大类,分别构成DNA和RNA的基本结构单位。所有的核苷酸都可在其5′位置连接一个以上的磷酸基团。从戊糖开始的第一、二、三个磷酸残基依次称为α、β、γ。α和β及β和γ之间的键是高能键,为许多细胞活动提供能量来源。核苷三磷酸缩写为NTP,核苷二磷酸缩写为NDP。5′核苷三磷酸是核酸合成的前体。

4. DNA 分子的一级结构

核酸是由一系列交替存在的糖和磷酸残基形成的骨架连接成的多聚核苷酸链(见图 2-4)。一个戊糖环的 5′位置通过 3′,5′-磷酸二酯键与另一个戊糖环的 3′位置相连,含氮碱基突出于骨架上。多聚核苷酸链一端的核苷酸具有自由的 5′磷酸基团;而另一端的核苷酸具有自由的 3′羟基基团。传统上按照 5′到 3′的方向书写核酸顺序,即左边为 5′末端,右边为 3′末端。

DNA 是由大量的 4 种脱氧核糖核苷酸通过 3′,5′-磷酸二酯键连接起来的直线或环形多聚体。所谓 DNA 的一级结构就是指 DNA 分子中的核苷酸排列顺序。生物的遗传信息通过核苷酸不同的排列顺序储存在 DNA 分子中,生物界物种的多样性即寓于 DNA 分子 4 种核苷酸千变万化的序列排列之中。

第三节　DNA 的二级结构[5-8]

一、DNA 双螺旋模型的提出

1953 年,Watson 和 Crick 两人提出了 DNA 双螺旋模型(double helix model)(图 2-5),这个模型不仅解释了当时所知道的 DNA 的一切理化性质,还将结构和功能联系起来,大大地推动了分子生物学的发展,具有划时代的意义。当时提出 DNA 双螺旋结构的根据主要有三方面:

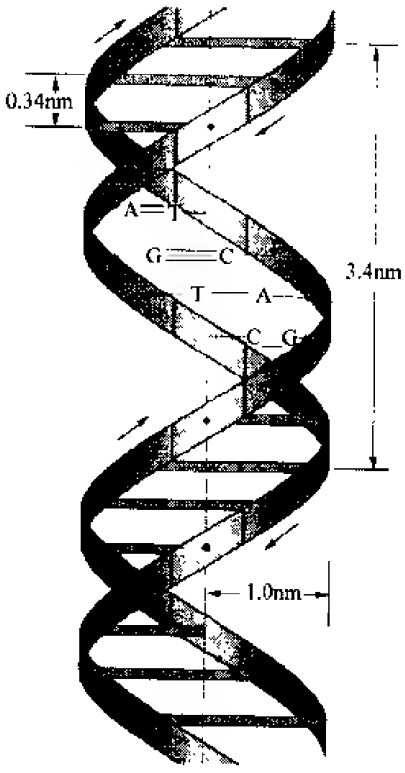


图 2-5　DNA 分子的双螺旋模型

1) 利用 X 射线衍射方法研究 DNA 纤维的结构,表明 DNA 分子可能具有规则的螺旋结构,每 3.4nm 形成一圈,直径为 2nm。因为邻近的核苷酸之间的距离为 0.34nm,所以每圈有 10 个核苷酸。

2) Chargaff 等发现的 DNA 中碱基含量 A=T,G=C 的定律。

3) DNA 的密度表明螺旋由两条多聚核苷酸链组成。螺旋的直径恒定,故每条链上的碱基只能朝向螺旋内部,嘌呤总与嘧啶相对。根据 4 种碱基的物化数据,碱基分成大小两种,嘌呤比嘧啶大些,若 A-T 配对或 G-C 配对,所形成的碱基对的几何大小接近;同时从嘌呤和嘧啶的氨基和酮基的键长和键角来看,A-T 间和 G-C 间能形成合适的氢键。

根据 Watson 和 Crick 所提出的模型,DNA 双螺旋结构具有以下特征:

(1)主链　脱氧核糖和磷酸基通过 3′,5′磷酸

二酯键连接形成螺旋链的骨架。两条主链以反平行的方式围绕同一中心轴相互缠绕，组成双螺旋。两条链均为右手螺旋。磷酸与核糖主链处于螺旋的外侧，在磷酸基团上带有负电荷。碱基处于螺旋的内侧。核糖平面与螺旋轴平行。

(2)碱基配对　两条核苷酸链依靠彼此碱基之间形成的氢键相连结合在一起，而且必须是嘌呤和嘧啶相配。但A和C，G和T之间不能形成适合的氢键，只有A和T，C和G配对才能保证形成正确的双螺旋结构。其中A与T配对，形成两个氢键；G与C互补配对，形成三个氢键，所以GC之间的连接较为稳定。上述碱基之间的配对原则称为碱基互补(见图2-6)。碱基互补原则具有极重要的生物学意义，DNA复制、转录、反转录的分子基础都是碱基互补。碱基配对的概念是核酸参与的所有过程的中心，碱基对的破坏是双螺旋分子发挥功能的关键特征，而形成碱基对的能力是单链核酸发挥功能所必不可少的。

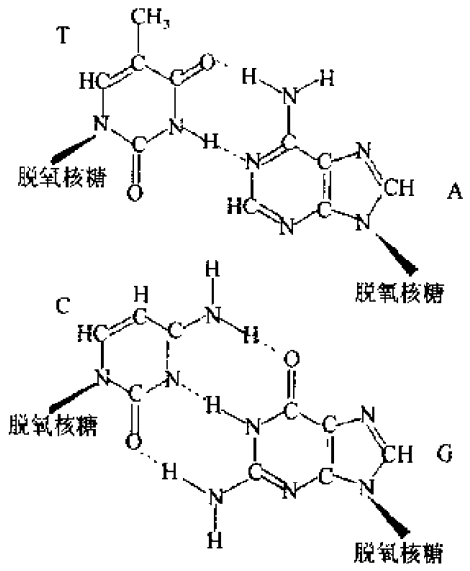


图2-6　DNA分子中的A-T、G-C碱基配对

(3)螺旋参数　碱基是平面的结构与螺旋轴垂直，两个相邻碱基对之间相距的高度，即碱基堆积距离为0.34nm，两个相邻碱基对之间绕螺旋轴旋转的夹角为36°，因此双螺旋链中的任意一条链绕轴一周的距离螺距为3.4nm，其中含10nt。双螺旋的平均直径为2nm。

(4)大沟和小沟　配对的碱基并不充满双螺旋空间，且碱基对占据的空间不对称。沿螺旋轴方向观察，双螺旋的表面形成两条沟，一条宽(2.2nm)，叫大沟(wide groove)，一条窄(1.2nm)，叫小沟(narrow groove)。大沟中碱基的差异很易被识别，是蛋白质结合特异DNA序列的位点，对于蛋白质识别DNA双螺旋结构上的特异信息非常重要。

二、维持DNA双螺旋的力

DNA双螺旋结构在生理状态下是很稳定的。维持这种稳定性的主要因素有以下几点：

(1)氢键　双螺旋中互补碱基对之间存在氢键(hydrogen bond)，G、C对之间有三条氢键，A、T对之间有两条氢键。由于G、C对比A、T对更稳定，DNA双螺旋结构的稳定性与与G+C的百分含量成正比。

(2)碱基堆积力　嘌呤与嘧啶碱呈疏水性，分布于双螺旋结构内侧。大量邻近碱基对的堆积，使双螺旋结构内部形成一个强大的疏水区，与介质的水分子隔开。

(3)其他作用因素　大量存在于DNA分子中的其他弱键在维持双螺旋结构的稳定上也起一定作用磷酸基团上的负电荷与介质中的阳离子之间组成的离子键，可以有效地屏蔽磷酸基之间的静电斥力，在体内的天然状态下，带正电荷的蛋白质可以中和负电荷；范德瓦耳斯力(van der Waals force)的大量累积，加强了碱基之间的疏水相互作用。

基因组的大多数DNA都以双螺旋B型DNA的形式存在，但不能把DNA视为具有

完全不变的双螺旋结构的碱基对序列。双螺旋结构是可变的，结构变化的能力是 DNA 发挥功能的关键，变化的范围从 B 型双螺旋参数的微小改变到双螺旋的分开而形成单链核酸。

三、双螺旋结构的基本形式

Watson 和 Crick 所用的资料来自相对湿度为 92% 时所得到的 DNA 钠盐纤维。这种 DNA 称为 B 型 DNA。很早人们就知道 DNA 可以以多种形式的双螺旋结构存在，如 A、B、C 等构象。这些结构都是右手螺旋，当溶液环境发生适当的改变时，这些构象之间可以发生相互转化。而且在每种构象家族中，参数会有轻微的变化。Z 构象是一种完全不同的左手螺旋结构（见图 2－7）。

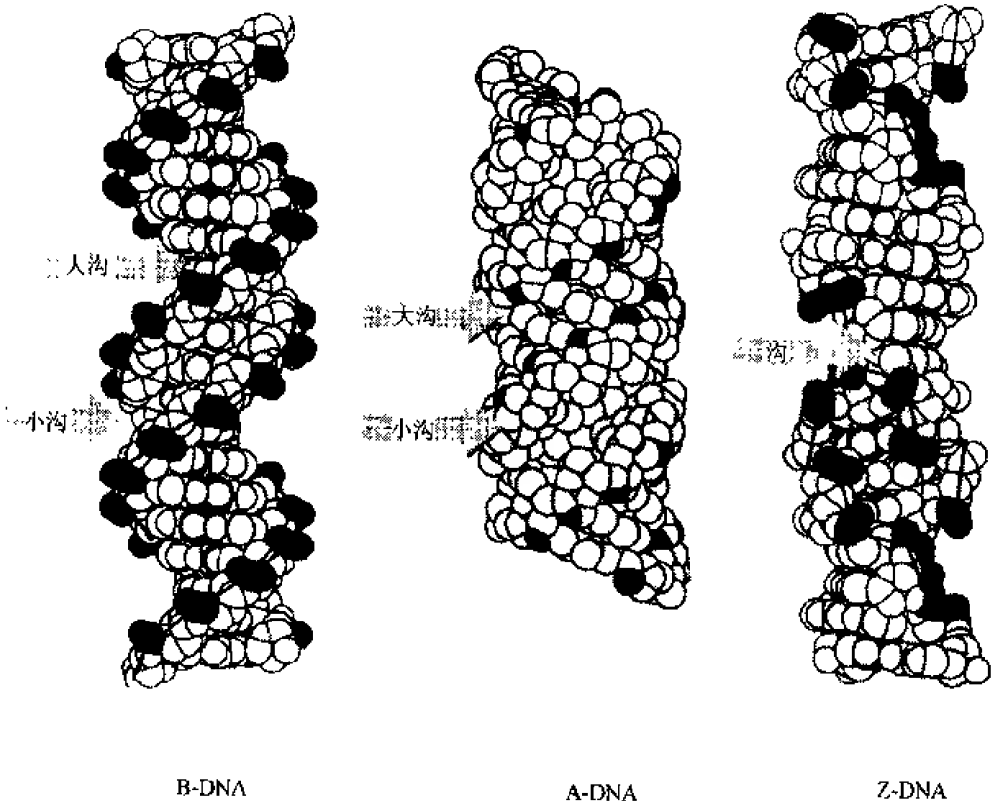


图 2－7　A、B、Z 型 DNA

1. A-DNA 的结构

相对湿度为 75% 以下获得的 DNA 纤维具有不同于 B-DNA 的结构特点，称为 A-DNA。DNA-RNA 杂交分子、RNA-RNA 双链结构均采取 A 构象。A-DNA 也是由反平行的两条多核苷酸链组成的右手双股螺旋。但是螺旋体较宽而短。每圈螺旋 11 个碱基对，螺旋扭角为 33°；每个碱基对的螺旋升值为 0.26nm；碱基倾角 13°，碱基平面不再与螺旋轴垂直，螺旋轴位于大沟中，小沟宽而浅，大沟很深，且磷酸基团覆盖其上，使其很难进入。

2. B-DNA 的结构

生物体内天然状态的 DNA 几乎都以 B-DNA 存在。自从其发现时起，描述 B 型双螺旋的一些参数已进行了调整。目前认为，每圈螺旋的平均碱基对数目为 10.4 而不是经典 B 型模型的每圈 10 个。邻近碱基对的螺旋扭角也改变为 34.6°。

这些数据是平均值，在 DNA 分子特殊的区域，如特殊的碱基序列和处于特殊环境的序列中这些数据可能会不同。因此我们取每圈为 10.4 代表 B 型 DNA 在生理状态下的平均值。在 B 型结构家族中，一段特殊的序列的参数值可能接近但并不一定完全等于平均值，其变化范围是 10.0 到 10.6。

3. Z-DNA 的结构

在已观察到的具有 Watson-Crick 碱基对的 DNA 双螺旋中，Z-DNA 具有某些不同的结构。A. Rich 等用 X 射线衍射法分析人工合成的 DNA 小片段的晶体时，出乎意料地发现，d(CGCGCG)是左旋的双螺旋。在这种双螺旋中两条六核苷酸分子链以反平行方向排列。在 d(CGCGCG)晶体上，磷酸核糖骨架呈 Z 字形走向。所以称为 Z-DNA。Z-DNA 并不只限于 GC 相间的 DNA 顺序中，在双螺旋多核苷酸链中，只要嘌呤和嘧啶碱基交替排列，在一定的条件下(如高盐浓度)，都有可能出现 Z 构象。对 Z-DNA 抗体的研究发现，天然 DNA 中的部分区域可以与 Z-DNA 的抗体结合，因此左手螺旋 DNA 可能存在于天然 DNA 的特殊区域中。在 B-DNA 向 Z-DNA 的转变过程中，鸟嘌呤绕糖苷键旋转 180°，导致核苷酸成顺式构象。而 C 核苷胞嘧啶和核糖整个旋转，使 C 核苷保持反式构象。这种 G 和 C 的不同构象引起了磷酸核糖骨架的锯齿形走向。

Z-DNA 为左手双螺旋；每圈螺旋 12 个碱基对，螺旋扭角为 −51°(G-C)和 −9°(C-G)；螺距 4.6nm，每个碱基对的螺旋上升值为 0.35nm(G-C)和 0.41nm(C-G)；碱基倾角 9°，碱基平面不再与螺旋轴垂直；螺旋轴位于小沟中，大沟已不存在，小沟窄而深，并具有更多的负电荷密度。表 2-1 归纳了 A、B 和 Z 型 DNA 螺旋的结构参数。

表 2-1 A、B、Z-DNA 的比较

	B-DNA	A-DNA	Z-DNA
每圈 bp 数	10.4	11	12
每 bp 转角(度)	34.6	32.7	−30.0
每 bp 上升距离(nm)	0.34	0.26	−0.38
螺旋直径(nm)	1.9	2.3	1.8

四、变性和复性

1. 变性

前文提到维持双螺旋的力由氢键和碱基的疏水相互作用提供，因此凡是破坏氢键和

疏水作用的因素都能导致双螺旋的破坏，如加热、极端的 pH、有机溶剂、低盐浓度等。DNA 双螺旋区的氢键断裂，使双螺旋的两条链完全分开变成单链，链分离的过程叫做变性(denaturation)或熔解(melting)。

当将 DNA 的稀盐溶液加热到 80～100℃时，双螺旋的结构即发生解体，两条链分开，形成无规则线团。DNA 变性的特点是爆发式的，变性作用发生在一个很窄的温度范围内，并导致一系列的理化性质发生剧烈变化，如溶液的黏度降低、沉降速度加大、浮力密度上升、紫外吸收值升高。我们可利用这些性质跟踪 DNA 的变性过程，其中最有用的变化是溶液光密度值的增加，这叫做增色效应(hyperchromic effect)，这是由于双链变性后碱基的充分暴露增加了在紫外区的光吸收。

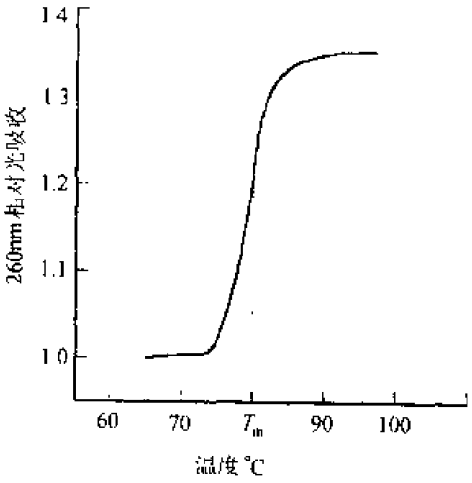


图 2-8 DNA 的熔解曲线

通常把 DNA 的双螺旋结构失去一半时的温度称为该 DNA 的熔点或熔解温度，用 T_m 表示。在适宜的生理条件下，DNA 的 T_m 值一般在 85～95℃之间，所以在细胞状态下，双螺旋 DNA 是稳定的结构。但当 DNA 复制或表达时，双螺旋链必须被分开。链分离是暂时的，在任意时刻基因组中只有很少的一部分区域呈单链状态。从双螺旋 DNA 的生物学功能上看，它的中心特性是能够不断裂共价键而分离双链，这使之能够以很高的速度在生理条件下进行链分离和重建，以维持其基本功能。

如图 2-8 所示，我们可用光吸收值的变化表示 DNA 的熔解曲线(melting curve)，曲线常常表现为相同的形状，但在温度坐标上的绝对位置(T_m)却受 DNA 的碱基组成和变性环境的影响。

2. 影响 DNA T_m 值大小的有关因素

(1)DNA 的均一性　均一性越高的样品，熔解过程越是发生在一个很小的温度范围内。

(2)G-C 的含量　G-C 含量越高，T_m 值越高，成正比关系。这是因为 G-C 比 A-T 碱基对更稳定的缘故。所以测定 T_m 值可推算出 G-C 的含量。其经验公式为：(G-C)% = (T_m - 69.3) × 2.44。

(3)介质中的离子强度　一般说离子强度较低的介质中，DNA 的熔解温度较低，熔解温度的范围也较窄。而在较高的离子强度的介质中，情况则相反。所以 DNA 制品应保存在较高浓度的缓冲液中或溶液中，常在 1 mol/L NaCl 中保存。

3. 复性及杂交

变性 DNA 在适当条件下，两条彼此分开的链又可重新缔合成为双螺旋结构，这个过程叫做复性(renaturation)。复性依赖于两条互补链之间专一性的碱基配对。复性反应经过两个阶段，首先在溶液中 DNA 的单链以随机发生相遇，如果它们的序列可以互补，两

条链就通过碱基配对形成一个短的双螺旋区域;然后这个碱基配对的区域延伸至整个分子,形成一个长的双螺旋分子。DNA复性后,许多理化性质又得到恢复。

复性是指变性分离的两个互补序列之间的反应,其实不同来源的两个互补核酸序列也能通过相互退火形成双螺旋结构,例如一条链来源于DNA,而另一条链来源于RNA,通常把这样的反应称为杂交(hybridization)。因为只有互补序列间能形成双螺旋结构,所以两种核酸之间杂交的能力可以准确反映它们之间的互补程度。核酸的杂交在分子生物学和分子遗传学的研究中占有十分重要的地位。杂交反应的原理是使两种核酸分子相互混合并测定形成的双链分子的数量。杂交反应通常有两种形式,液相杂交和滤膜杂交。两种并不完全互补的序列也能杂交,如果两个序列十分相似但不相同,它们可形成不完整的双螺旋,在不能互补的位置碱基配对会发生中断。

很多因素影响复性的速度:①简单序列DNA分子,彼此发现互补序列并不困难,因此复性很快。而复杂序列的DNA在与其互补链碰撞之前,必然会遇到很多非互补链。因此在同样条件下,序列复杂的DNA比序列简单的DNA复性所需的时间要长。②同一种DNA浓度高时,互补序列碰撞的机会增加,复性速度也快。③DNA片段的大小也影响复性的速度,因为大片段的扩散速度慢,减少了发现互补序列的机会,因此在复性实验中,往往把DNA链剪切成一定大小的片段。④温度也影响复性的速度,低温减少了互补链碰撞的机会,而且已经误配的片段解开以寻找正确的互补链在低温下也很困难。因此复性温度不能过低,一般比熔点低25℃左右是最佳温度。⑤溶液的离子强度也影响复性速度,两个带同样电荷的DNA会互相排斥,因此在复性溶液中必须有一定浓度的盐。

第四节　二级结构的其他形式[9~11]

一、单链核酸形成的二级结构

尽管DNA通常以双螺旋形式存在,但一些病毒的基因组却是由单链DNA组成的,在细胞中还有很多其他形式的核酸,RNA经常以单链形式存在。单链核酸也可能产生双螺旋区域。如果单链分子中的一些序列可以与另一些进行碱基配对,就可产生单链分子内的双螺旋区域。而且一个单链分子可以与另一个独立的互补单链分子碱基配对而形成一个分子间的双螺旋。互补单链之间的碱基配对并不仅局限于DNA-DNA或RNA-RNA的相互作用,也可能发生于DNA和RNA分子之间。

因为RNA是主要的单链核酸,所以通常用RNA来考察单链核酸形成的双螺旋区域,所使用的方法也同样适用于单链DNA。RNA的一级结构与DNA一致:都是由5′-3′核糖磷酸键连接的多聚核苷酸。由于是单链核酸,RNA分子在空间中的折叠采取随机方式,但分子中存在部分序列之间的碱基配对能使一个区域和另一个区域保持固定的位置。

当同一个RNA分子的一段碱基序列附近紧接着一段它的互补序列时,核酸链有可能自身回折配对产生一个反平行的双螺旋结构,叫做发夹(hairpin)。发夹结构由叫做茎的碱基互补配对的双螺旋区域和末端不配对的环构成。当互补序列在分子中距离相对较远时,它们接近并形成双链区域时将产生一个很长的单链环。

目前无法准确预测特定单链核酸序列形成的二级结构,由核酸分子的物理性质可推

断分子碱基配对的整体情况，但不能表明是哪个具体的区域参与配对。核酸分子的单链和双链区域对一些核酸酶有不同的敏感性，因此我们能够分析区别哪些特殊的区域参与了碱基配对，但通常这些技术只适用于相对短的分子。

因为在一个单链分子中可能有很多区域可以与一个特定区域发生互补配对，所以为了预测在一个单链核酸序列中到底会出现何种二级结构，必须确定实际发生配对的可能。当存在多种可能的结构时，人们利用一些方法来预测这些结构之间相对的稳定性。当然这些方法把 RNA 视为独立存在的结构，而忽略了其他因子（如结合蛋白）对核酸结构的影响。

这些方法的基础是计算形成每种结构的自由能变化，碱基配对结构的形成是释放能量的过程，结构的稳定性是由所释放能量的多少决定的。因此当两个结构产生的自由能不同时，具有较低自由能的结构更可能形成。

形成一个双链结构总的自由能与下列因素有关：

1）起始双螺旋的自由能，用来产生第一个碱基配对，它是一个正值（14.2kJ/mol），表明这是一个耗能过程；

2）碱基互补配对所释放的自由能的总和，是一个负值；

3）形成双螺旋的序列有时不能完全互补配对，保持这种不配对的碱基状态需要能量，它的总和是一个正值。

总的自由能是三种能量的相加，为了形成一个双螺旋反应的总能量必须是一个负值，即碱基配对反应释放的能量总和必须明显超过起始双螺旋及克服不配对的区域需要的能量总和。

为了计算形成一个碱基对释放的能量，我们必须考虑碱基对内部的作用及相邻碱基对的影响。G：C 之间的碱基对具有三个氢键，比具有两个氢键的 A:U 配对更稳定；碱基对间堆积的自由能取决于相邻的含氮碱基的组成，所以每个相邻二核苷酸对组合的自由能是不同的。因此每个反应释放的自由能的精确数值取决于各碱基对的综合作用，这意味着对每个可能的二核苷酸对的自由能值受碱基组成和碱基排列的双重影响。形成双螺旋区域的自由能是每个相邻的碱基对组成的二核苷酸对的自由能的和。

表 2-2　相邻二核苷酸对的自由能（kJ/mol）

5′ \ 3′	A	U	G	C
A	-3.8	-3.8	-7.1	-8.8
U	-4.6	-3.8	-7.5	-9.6
G	-9.6	-8.8	-12.1	-14.2
C	-7.5	-7.1	-8.4	-12.1

在 RNA 中还可形成另一种不规则的碱基对 G·U 配对，而在单链 DNA 中的双链区域也能形成 G·T 配对。G·U 配对的自由能变化的，具体数值取决于相邻的碱基对（表 2-2）。与具有完整的双螺旋结构的双链 DNA 不同，RNA 中的双链区域通常由两个不完全互补的单链序列形成，双螺旋区域完整性的中断表现为下列形式：①当双螺旋由邻近的互补序列回折配对形成时，在双螺旋区域的末端产生发夹环（hairpin loop）结构；②在可能的互补区域内具有不能碱基配对的相应区域时将产生内部环（interior loop）；③当两个可能的互补序列中的一个包含一段不配对的多余序列时产生凸环（bulge loop）。

所有这些不配对的区域都阻碍双螺旋的形成，为了使这些区域保持特殊的结构必须提供能量，这将抵销碱基配对和堆积释放的自由能，所以在自由能的计算时必须考虑它们的影响。

二、反向重复与二级结构

前文讲到，当一个单链核酸包含两段互补序列时，能通过碱基配对形成发夹结构。双链 DNA 中也有与之相对应的结构，它由以相反方向排列的完全相同的序列组成，这种结构叫做反向重复(inverted repeat)。

反向重复中的序列叫做回文序列(palindrome)。回文序列是双螺旋 DNA 中的一段序列，如果按确定的方向读双链中的每一条链的序列都是相同的。序列$\begin{matrix}5'GGTACC3'\\3'CCATGG5'\end{matrix}$就是回文结构，因为每条链从 5′→3′方向读的顺序，都是 GGTACC。

回文序列是二元对称结构，对称轴把反向重复分开。在对称轴的每侧都有以相反方向存在的相同序列 GGT/CCA。反向重复的两个拷贝不一定完全相同，下面的一种情况也是反向重复：$\begin{matrix}GGTNNACC\\CCANNTGG\end{matrix}$。反向重复可形成单链或双链形式的二级结构。

在双链 DNA 中，如果两条互补链分开，每条链上的互补序列都有机会发生碱基配对而形成一个发夹。两个相对的发夹结构形成了一个十字形结构(cruciform)，对应于 4 个双螺旋区域的交叉点。原来的双螺旋位于十字形结构的两侧，单链形成突出的发夹。

对于在体内是否能形成十字结构是有争议的。双链 DNA 似乎不会放弃连续的双螺旋结构而自发地产生一个突出的发夹。双链分开需要较多的能量(209.3kJ/mol)，而且在形成十字结构时不能重新获得全部损失的能量，所以十字结构与连续的双螺旋结构相比较不稳定(△G=75.4kJ/mol)。但在体外适宜的条件下可以形成十字结构(图 2-9)。

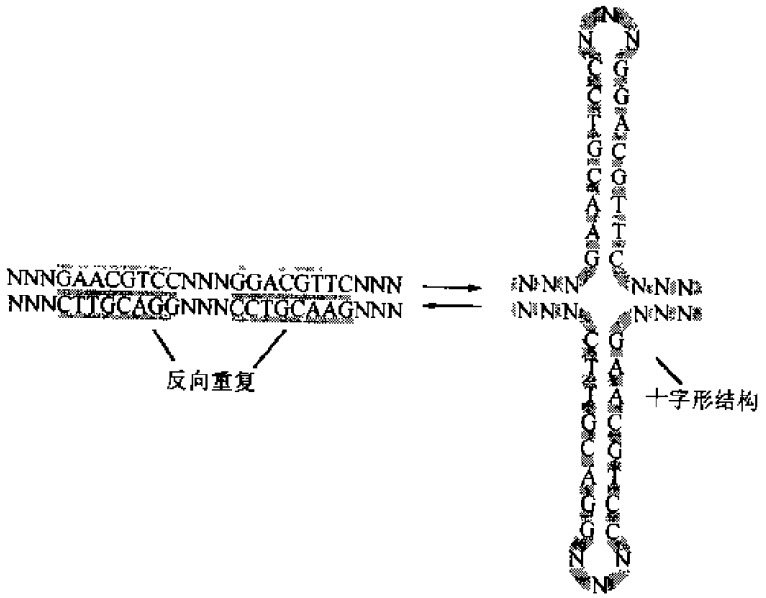


图 2-9 回文结构转化为发夹结构及十字形结构

三、DNA 四链结构

特殊形式的 DNA 结构还有四链结构(tetrasomy strucure),在此结构中基本结构单元是鸟嘌呤四联体(G-quartet)。人们早就观察到各种鸟苷酸在一定条件下形成凝胶的现象,其原因就是鸟嘌呤之间的特异相互作用导致 G-四联体的形成。在 G-四联体结构中,4 个 G 有序地排列在一个正方形片层中,相邻碱基之间以非正常的 G-G 氢键相连,形成首尾相接的环形结构。其中每个鸟嘌呤碱基都是氢键的供体和受体。片层中央是由电负性的羰基氧形成的口袋,可以容纳一价阳离子,并与之相互作用。

后来,研究人员利用 X 射线纤维衍射技术,表明多聚鸟苷酸采取的是四螺旋 DNA 结构,其基本结构单元就是 G-四联体。整个四链结构可看成由多个 G-四联体片层以螺旋方式堆积而成(见图 2-10)。其中每一片层包含 4 个鸟嘌呤碱基,分别来自 4 条聚鸟苷酸链。在此结构中螺旋扭角为 30°,每个片层沿螺旋上开 0.34nm。G 的糖苷链为反式构象。磷酸骨架均采取平行的 5′-3′排列方式。

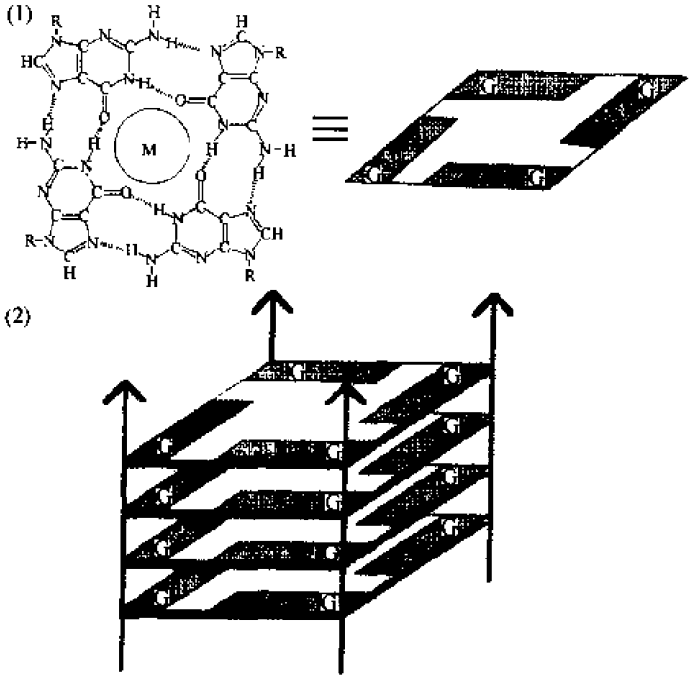


图 2-10 鸟嘌呤四联体结构及其构成的 DNA 四螺旋结构
(1)鸟嘌呤四联体的结构,4 个鸟嘌呤碱基呈环状排列,每个碱基都是氢键的供体和受体。在鸟嘌呤四联体的中心是一价阳离子;(2)鸟嘌呤四联体的堆积形成了 DNA 四螺旋结构。

从 poly-G 结构可以看出,串联重复的鸟苷酸是形成四链 DNA 的必要条件。真核生物染色体的端粒 DNA 中,有许多富含鸟嘌呤的串联重复(tandem repeat)序列,如尖毛虫的端粒序列为 $d(TTTTGGGG)_n$,四膜虫为 $d(TTGGGG)_n$,人类为 $d(TTAGGG)_n$,这些重复序列可以写成通式 $d[T_{1-3}-(T/A)-G_{3-4}]_n$,这些富含 G 的序列总是位于端粒的 3′末端,相应的互补 5′末端则富含 C。因此端粒在一定条件下有可能采取与 poly G 类似的,以

G-四联体为基本单位的四链 DNA 结构。实验表明,G-四联体的多种特征,在端粒 DNA 结构中也可以观察到。

1) 在非变性电泳中,端粒 DNA 序列表现出很高的泳动性,暗示其含有紧密压缩的结构。

2) 与 G-四联体一样,端粒 DNA 特殊结构的形成需要一价阳离子的存在。

3) 在一定条件下,端粒 DNA 表现出与 G-四联体相同的对单链核酸酶水解的抗性。

4) 核磁共振(NMR)及 X-射线衍射技术表明在其结构中存在 G-G 氢键交联。

因此,端粒 DNA 可以形成以 G-四联体为基本单位的四链核酸结构。与 poly G 相比,端粒 DNA 还表现出一些新的特征。在端粒 DNA 形成四链结构的过程中,由于串联重复的 G 相互配对形成 G-四联体,将导致 DNA 链回折,含 A/T 序列的部位产生环状结构,分布在 G-四联体堆积的上下两端。由于回折,4 条链的磷酸骨架会出现反平行的 5′-3′方向排列。从而产生反常的顺式鸟嘌呤糖苷键。这种构型,也存在于 Z-DNA 中。不同的磷酸骨架方向,顺-反糖苷键在 G-四联体中的分布,以及环的连接方式,将导致多种结构异构体的产生(图 2-11)。

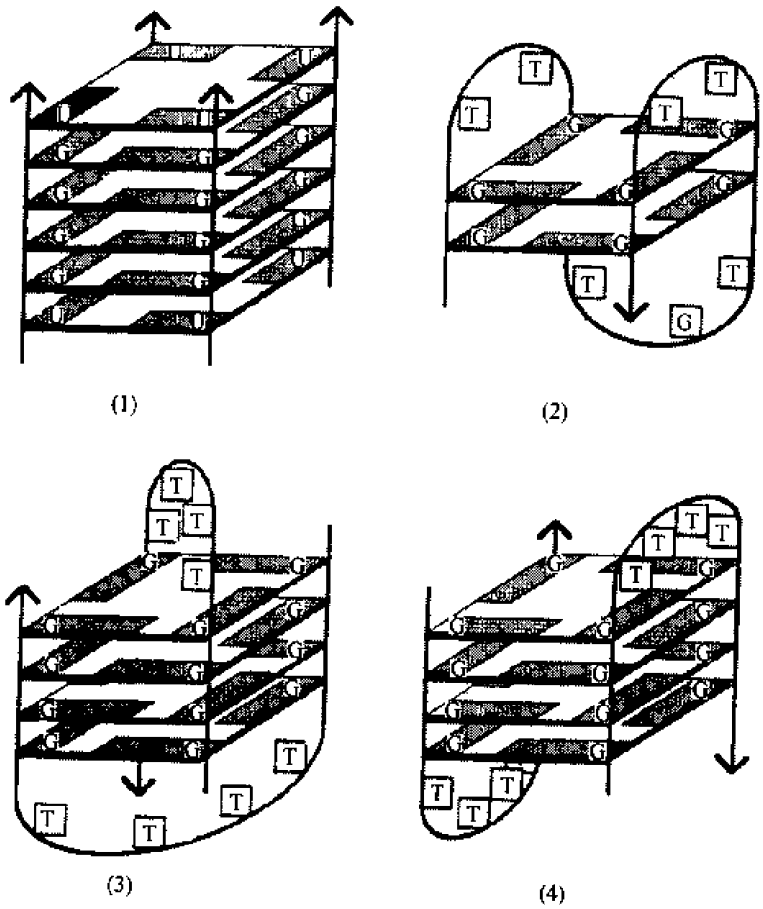


图 2-11 鸟嘌呤四联体结构的不同形式

(1)$(UGGGGU)_4$ 形成的结构,由 4 条链构成;(2)d(GGTTGGTGTGGTTGG)结合凝血酶时形成的结构,由单链构成;(3)d(GGGGTTTTGGGG)在 Na^+ 溶液中的结构,由两条链构成;(4)d(GGGGTTTTGGGG)在 K^+ 溶液中的结构。

除端粒 DNA 以外，其他富含 G 的 DNA 序列也可能产生以 G-四联体为基础的结构。包括：免疫球蛋白铰链区基因中富含 G 的部位、成视网膜细胞瘤敏感性基因和 tRNA、*SupF* 基因上的一些特殊序列等。任何序列，只要有成串的串联重复的 G，都可能形成这种特异的结构。尽管这些序列在功能上与端粒 DNA 序列无关，但是它们的基本结构是类似的。

迄今为止，没有证据表明这种以 G-四联体为基础的核酸结构有什么生物学功能，但一些间接的实验表明这种结构存在于生物体内：

1）在酵母中观察到了由于端粒与端粒相互作用，导致线性质粒以环形形式存在。

2）发现尖毛虫端粒结合蛋白的 B 亚基可以催化 G-四联体的形成。而当 B 亚基出现点突变时，催化效率大大降低。生物体内这种催化功能的存在表明 G-四联体为基础的结构存在于生物体中。

3）在酵母抽提液中发现了以 G 四联体结构为底物的核酸酶。这种酶表现出结构专一性而不是序列专一性，暗示了 G-四联体结构在体内的存在。

四、三螺旋 DNA

早在 1957 年，Felsenfeld 等人就发现，当双链核酸的一条链都为嘌呤核苷酸（全嘌呤链），另一条链都为嘧啶核苷酸（全嘧啶链）时，如 polyA/polyU 、polydA/polydT poly dAG/poly d(CT)，就会发生转化形成核酸三链结构（triplex）。这个结构由一条嘌呤链和两条嘧啶链构成，但这种现象没有引起人们的广泛重视。

80 年代中后期，研究者发现，在超螺旋质粒中的一段全嘌呤/全嘧啶序列表现出一种特殊的 DNA 结构：H-DNA。其主要特征为 DNA 三螺旋。另外还发现，双螺旋 DNA 上的全嘌呤/全嘧啶序列可与第三条链（一条全嘧啶链或一条富含嘌呤的寡核苷酸）形成序列专一性的稳定复合物。这些复合物并不是由于单链 DNA 简单置换双螺旋结构中的一条链而形成的 D-loop 结构，而表现出三螺旋 DNA 的特征。这些发现促进了三螺旋 DNA 的研究。

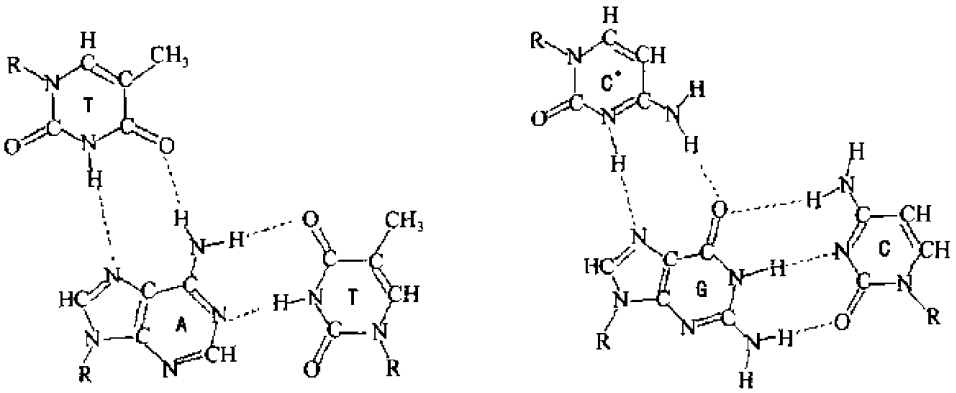
三螺旋 DNA 中三股链上的碱基都参与氢键的形成，它的基本结构单元是碱基三联体（triad），三个碱基分别来自三股链的相应部位。有多种配对形式。在由两条嘧啶链与一条嘌呤链组成的三螺旋中（YR * Y 型），有 CG * C，TA * T 两种三联体。在由两条嘌呤链与一条嘧啶链组成的三螺旋中（YR * R 型），有 CG * G，TA * A，TA * T，CG * A 4 种碱基三联体。（图 2-12）* 号左边的两个碱基形成正常的 Watson-Crick 碱基配对，* 号两侧的碱基则形成异常的碱基配对（Hoogsten 氢键），将三个碱基通过氢键连成三联体结构。中间的嘌呤碱基既参与双螺旋的形成，又参与第三链的结合，对三联体结构的形成有重大贡献。

三螺旋 DNA 就是由许多碱基三联体沿螺旋轴方向堆积而成。已发现的组成形式有多种。（图 2-13）这些结构表现出一些共同的基本特征：

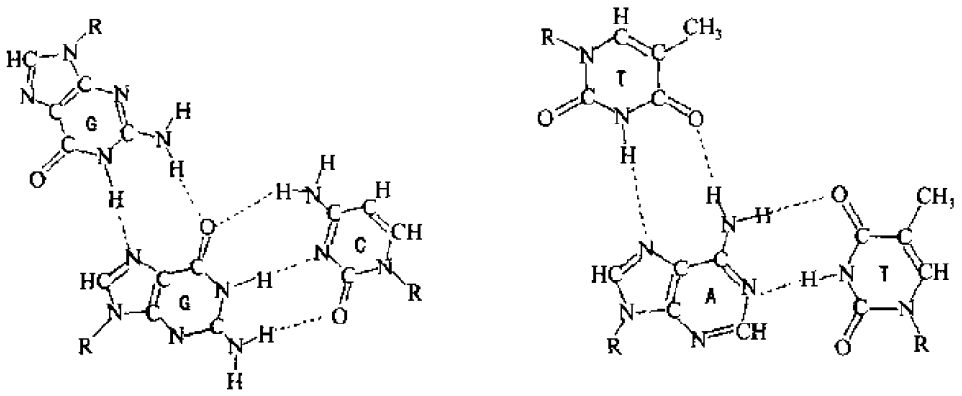
1）在三链形成过程中，双螺旋中的全嘌呤链与第三条链形成 Hoogsten 氢键配对；

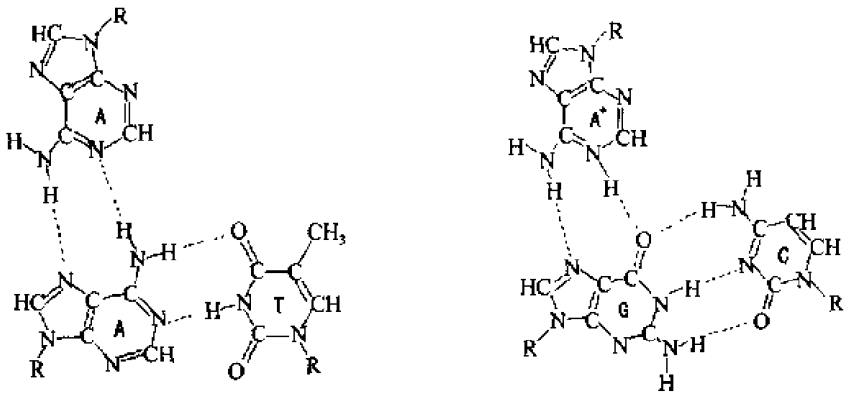
2）第三条链位于双螺旋结构的大沟中；

3）三链中化学性质相同的两条链采用反平行方式排列；



YR*Y碱基三联体：TA*T和CG*C+





YR*R碱基三联体：CG*G, TA*A, TA*T,和CG*A+

图 2-12 碱基三联体

4）三螺旋中的双螺旋的构型与 B-DNA 类似。

除了基本性质相同外，YR＊Y 型与 YR＊R 型三螺旋还有一些不同：YR＊Y 型在酸性 PH 下稳定，而 YR＊R 型的稳定性依赖于二价阳离子的存在，YR＊Y 型的三联体沿轴

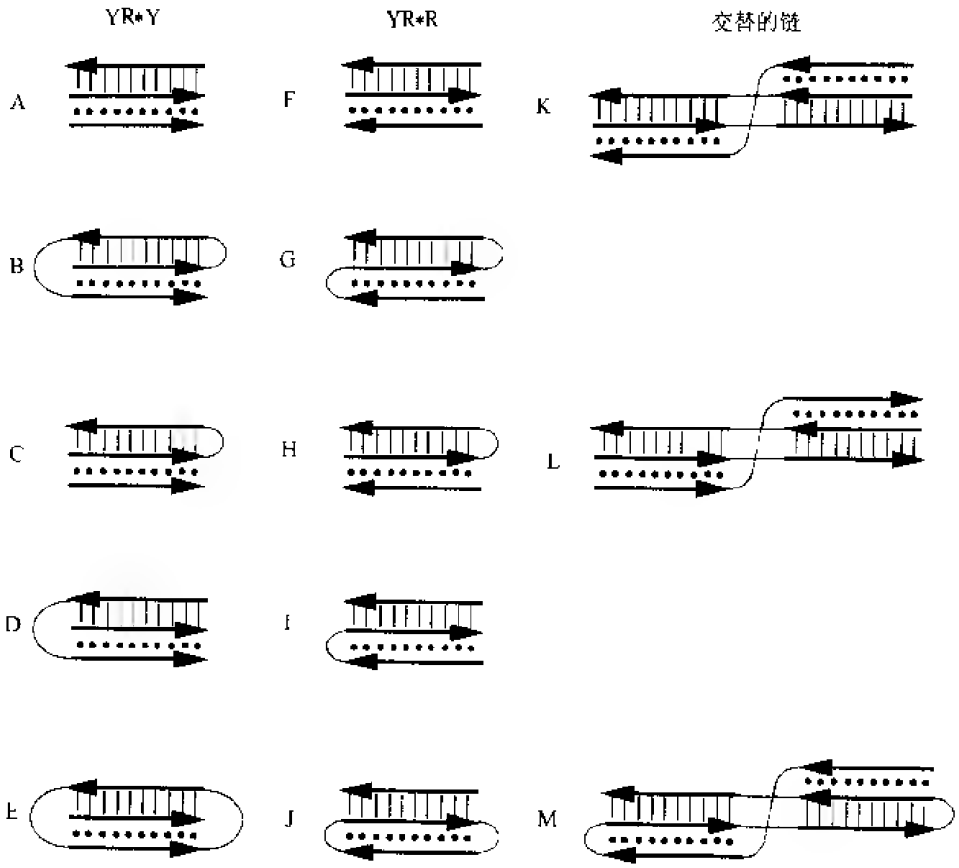


图 2-13　三螺旋 DNA

向上升 0.34nm，螺旋扭角 31°，YR＊R 型沿轴向上升 0.36nm，螺旋扭角 30°。

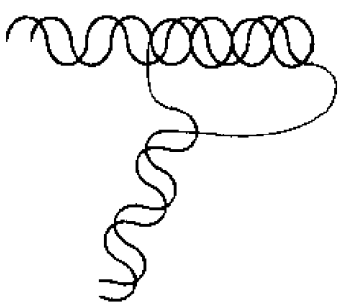


图 2-14　H-DNA 的结构

在研究海胆组蛋白基因间的间隔序列时，发现其中的 $d(GA)_{16}$序列对 S1 核酸酶敏感，暗示可能存在一种反常的 DNA 结构。此结构是依赖 pH 的，在酸性条件下稳定，而在中性 pH 下，则观察不到这种结构。因为 H^+ 能稳定这个结构，所以此结构被称为 H-DNA。

H-DNA 是分子内三螺旋 DNA，由一条嘧啶链和半条嘌呤链组成的 YR＊Y 型结构，剩下的半条嘌呤链以单链形式存在，这就解释了 H-DNA 对 S1 核酸酶的敏感性。图 2-14 中三螺旋的基本单位是 TA＊T、$CG*C^+$ 三联体。其中 C 的质子化对形成 $CG*C^+$ 三联体是必不可少的，这就是 H-DNA 的 pH 依赖性的原因。同时，H-DNA 的结构模型也要求每条链上的核酸序列必须是镜像重复序列。

分子内的 YR＊R 型三螺旋则叫作＊H-DNA。由于采取了多种三联体配对形式，所以三链结构的形成不需要 DNA 序列严格镜像重复。＊H-DNA 的稳定性依赖于二价阳离子存在，其原因还不清楚。

真核生物基因组中有大量可形成 H-DNA 的序列。但如上所述，YR * Y 三螺旋的形成需要在酸性 pH 条件，而 YR * R 三螺旋需要毫摩尔级的二价阳离子的存在。细胞中并不存在这样的条件，然而生物体内也具有促进三螺旋结构形成的因素：

1）研究表明，多胺类如精胺、亚精胺在生理 pH 下有利于 YR * Y 结构的稳定。这是由于结合多胺后，克服了三螺旋中相对高的负电荷，降低了磷酸骨架间的斥力。而在真核生物的细胞核中有毫摩尔级的多胺，为三链结构在体内存在提供了可能性。

2）增加 DNA 的负超螺旋密度，可促使三螺旋的形成。因为 H-DNA、* H-DNA 的形成降低了 DNA 超螺旋的扭曲张力。细胞核中的拓扑异构酶、旋转酶都会促使负超螺旋的增加。在转录过程中 RNA 聚合酶催化的聚合反应也会产生超螺旋微区。

3）DNA 合成时，需要中性 pH 及较高的 Mg^{2+} 浓度，有利于 YR * R 三螺旋的产生。

利用抗三螺旋 DNA 的抗体与真核染色体相互作用，表明在真核细胞中确实存在三螺旋 DNA 结构。

实验观察到三螺旋结构可阻止 DNA 的体外合成，所以有假说认为 H-DNA 参与了 DNA 的复制过程。可能的机制之一是：当 DNA 聚合酶到达核酸序列镜像重复的中央时，模板会回折，并与新合成的 DNA 链形成稳定的三螺旋，使 DNA 聚合酶无法延伸，从而终止复制过程。这个假说能否代表细胞内部的真实情况，还需要进一步的研究工作来证实。

第五节　DNA 的超螺旋结构[12]

一、超螺旋结构

在谈到 DNA 的双螺旋模型时，我们把 DNA 视为线性分子，但是体内的 DNA 常常是缺少自由末端的闭合结构。绝大多数原核生物的 DNA 都是共价闭合环，而真核生物的染色体虽为线性分子，但其 DNA 与蛋白质相互结合，以许多大环的形式存在，每个环的根部结合在一起形成类似环的结构。这种闭合环的存在使双螺旋上出现更多的结构限制。在 DNA 双螺旋中，大约每 10 个核苷酸长度旋转一圈，这时双螺旋处于能量最低的状态。如果这种正常的双螺旋 DNA 额外地多转几圈或少转几圈，就会使双螺旋中存在额外的张力。如果双螺旋的末端是开放的，这种张力可以通过链的转动而释放出来，使 DNA 恢复正常的双螺旋状态。但在共价闭合环状 DNA 或与蛋白质结合的 DNA 中，DNA 分子两条链不能自由转动，这些额外的张力就不能释放到分子之外，导致 DNA 分子扭曲以缓解这种张力，DNA 分子的这种扭曲称为超螺旋(supercoil)(见图 2－15)。一个闭合分子必须在 DNA 的每条链上都没有断裂，即使在一条链上出现的断裂都有可能导致超螺旋的消失。一个分子无论是闭合还是开放结构，只要缺少超螺旋结构，就被称为松弛结构。超螺旋 DNA 是一种常见的三级结构类型，生物体中绝大多数 DNA 是以超螺旋的形式存在的。

超螺旋是有方向的，有正超螺旋和负超螺旋两种。负超螺旋使 DNA 绕其轴以与右手双螺旋的顺时针方向相反的方向缠绕，通过减少每个碱基对的旋转松弛双链的互相缠

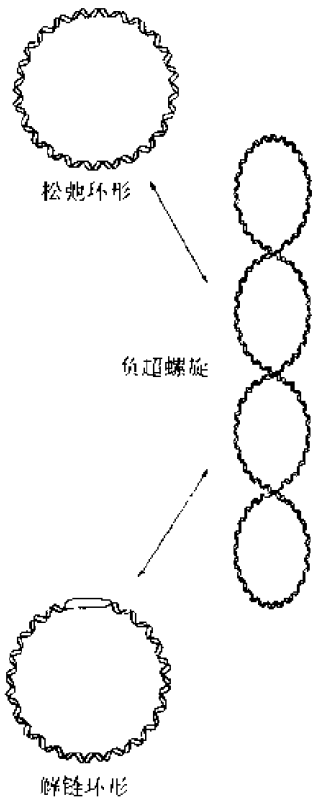


图 2-15 环状 DNA 的不同构象

绕，使 DNA 分子减轻了因自身双螺旋而产生的扭曲张力，因此负超螺旋的作用是解旋。但如果过度扭转，将会导致部分碱基对的破坏；如果 DNA 按照其内部双螺旋缠绕方向相同的方向缠绕形成超螺旋，就会产生相反的结果。正超螺旋的引入导致绕双螺旋的张力更大，使结构更加紧张，因此正超螺旋是过渡缠绕的超螺旋。体外实验可以产生这种正超螺旋。负超螺旋把扭曲张力传递到 DNA 上，使 DNA 解旋，而正超螺旋使双螺旋缠绕得更紧。超螺旋使环状双链结构更紧密，因而在离心时沉降更快。

下面利用一些拓扑学的概念说明超螺旋的结构：

(1) 连接数(linking number)　连接数是指双螺旋 DNA 中，两条链互相交叉的次数，以字母 L 表示。它是一个整数。一个闭合环状 DNA 分子，只要其主链的共价键不发生断裂，连接数是不会改变的。而且 L 值相同的 DNA 之间可以不经过链的断裂而互相转变。

(2) 扭转数(twisting number)　代表 DNA 分子一条链绕另一条链的扭转数，即 DNA 分子中双螺旋的数目，是双螺旋自身的性质，以 T 表示。

(3) 缠绕数(writhing number)　指双螺旋分子在空间上相对于双螺旋轴的扭曲，也等于超螺旋数目。以 W 表示。

L、T、W 三者的关系为：$L = T + W$。T 与 W 值可以是小数，但 L 值必须是整数。对于松弛的分子，$W = 0$，而 L 与 T 相等。当拓扑异构体之间的 L 值相差为 1，就可以通过琼脂糖凝胶电泳将它们分开。

(4) 比连接差(specific linking difference)　比连接差以 λ 表示，用来表示 DNA 超螺旋的程度。$\lambda = (L - L_0)/L_0$。其中，L_0 是指松弛环形 DNA 的 L 值。如果 L 的变化都是由 W 的变化引起的(即 $\triangle T = 0$)，也可以将 λ 值视为超螺旋密度。大多数生物的超螺旋密度为 -0.05 左右。

二、超螺旋影响双螺旋的结构

在体内，DNA 的结构变化并不能自发产生，它们的发生需要处于特定环境，以使 DNA 受到多种约束。其中一些变化是由负超螺旋(negative supercoil)推动的。

负超螺旋的引入需要能量，所以超螺旋的存在使分子具有更多的能量，因此我们可把超螺旋看成能量的储存形式。负超螺旋能控制 DNA 结构变化的平衡，这使超螺旋 DNA 能实现松弛的 DNA 不能完成的结构转化。通常基因组表现为一定程度的负超螺旋，在体内一般为每 200 个碱基对一个负超螺旋，即超螺旋密度(superhelix density)为 -0.05，这相当于分子具有 37674.9 J/mol 的能量。在体内保持一定的负超螺旋水平是必要的，但并不是整个基因组中的超螺旋密度都相同，而是呈波动状态。可能 DNA 的转化只需要在特殊区域具有增加的超螺旋。

转化形式之一是链分离，它是由负超螺旋推动的。负超螺旋分子具有链分离所需要的额外的能量。解旋DNA需要的能量大小与碱基序列有关，事实上是形成双螺旋所释放能量的逆转。每分离10个碱基对大约需要54.2～209.3kJ/mol，所以超螺旋对应的能量只能够促使极少数碱基对的解旋，但其作用是很重要的，例如在复制DNA时起始双链分离。

三、拓扑异构酶

DNA拓扑异构体之间的转变是通过拓扑异构酶(topoisomerase)来实现的。这种酶可以改变DNA拓扑异构体的 L 值。DNA拓扑异构酶是催化DNA环的拓扑异构体之间转化的酶，它可催化一个DNA单链或双螺旋链穿过另一个单(双)链。

DNA拓扑异构酶可以分为两类：I型拓扑异构酶和II型拓扑异构酶。I型拓扑异构酶催化DNA链断裂和重新连接，每次只作用于一条链，即催化瞬时的单链断裂与连接，它们不需要能量辅因子ATP或NAD。典型代表为大肠杆菌DNA拓扑异构酶I和鼠DNA拓扑异构酶I。除了断裂单链的性质之外，两者的结构和性质都不同，分别属于IA和IB亚家族。IA型拓扑异构酶还有大肠杆菌DNA拓扑异构酶III和古细菌的反旋转酶，IB型拓扑异构酶则包括其他真核生物的DNA拓扑异构酶I。II型拓扑异构酶能同时断裂并连接双股DNA链，它们通常需要能量辅因子ATP。典型代表为大肠杆菌的DNA旋转酶(gyrase)，这个酶又叫做DNA拓扑异构酶II。此外还有大肠杆菌DNA拓扑异构酶IV和噬菌体(如T4)、真核生物(如酵母、果蝇等)的拓扑异构酶II。

1.大肠杆菌DNA拓扑异构酶I

大肠杆菌DNA拓扑异构酶I是最早发现的拓扑异构酶，它可催化负超螺旋DNA转化为松弛型DNA。当拓扑异构酶I结合负超螺旋DNA时，其结合能可克服短的DNA双螺旋解链的自由能，因此不需要ATP。但在松弛正超螺旋底物时，结合能不够，因此不能松弛正超螺旋DNA。拓扑异构酶I结合到单双链的接合部，催化单链区的DNA暂时断裂。在酶的催化口袋中，存在单链DNA结合位点。酶连接断裂位点处DNA的5′和3′端，形成了一个桥式结构的酶-DNA共价中间体，由酶结合的两端足够宽使另一条链能够顺利通过，然后两端又重新结合，完成了一次拓扑变化。其最小底物为六核苷酸，断裂位点是六核苷酸5′端的第4个核苷酸处。

IA型拓扑异构酶具有多个DNA结合位点，可结合双链或单链DNA。在酶催化口袋外的位点也参与DNA的结合或DNA的转移。大肠杆菌DNA拓扑异构酶I的C末端121个残基可结合DNA并提高酶-DNA复合物的稳定性。其598～737残基的缺失虽不影响单链寡核苷酸的断裂活性，但却不能使负超螺旋松弛，这个区域能结合Zn离子，是松弛活性所必需的。

2.真核生物DNA拓扑异构酶I

真核生物DNA拓扑异构酶I可结合在15～19个核苷酸的双链区域，其断裂位点在这段区域中间。真核DNA拓扑异构酶I倾向于结合超螺旋而不是结合松弛DNA。此酶

对正超螺旋和负超螺旋有相同的松弛能力，表明 IB 型拓扑异构酶识别的是 DNA 链的弯曲程度(flexural)而不是链的扭曲张力(torsional)，此酶可区分长度数千碱基对而连环数只相差一个的 DNA 异构体。

不能确定 IA 型拓扑异构酶的酶-桥模型在 IB 型拓扑异构酶中是否适用。IB 型拓扑异构酶可能采取 DNA 旋转模型进行 DNA 单链的转移，在 DNA 链通过缺口时酶随着与其相连的 DNA 链被动地旋转实现磁力线扭曲的松旋。在酶-桥模型中每次 DNA 断裂-重接循环只能去除一个超螺旋，而在 DNA 旋转模型的一个循环中所有的超螺旋都可被去除。但两种模式不是完全不同的，实际上无法确定拓扑异构酶采取何种模式。

3. 大肠杆菌拓扑异构酶 II

在大肠杆菌中起主要作用的拓扑异构酶是拓扑异构酶 II，又称为 DNA 旋转酶(DNA gyrase)，它由两个 A 亚基和两个 B 亚基组成。旋转酶 B 亚基中存在 ATP 酶区域，可以结合并水解 ATP。在 B 亚基存在下，A 亚基有超螺旋活性，其中氨基酸残基 1～523 有松弛超螺旋的活性，572～875 有结合 DNA 的能力。拓扑异构酶 II 能产生负超螺旋以抵销复制叉移动所产生的正超螺旋，无 ATP 时大肠杆菌旋转酶只能很慢地松弛负超螺旋但不能松弛正超螺旋。

细菌旋转酶对 DNA 的结合顺序是正超螺旋＞松弛＞负超螺旋，DNA 片段在酶上呈右手缠绕，围绕酶有 140bp 的片段。旋转酶的催化机制包括两个连续的转酯反应。在第一个反应中，酶活性部位的 Tyr 残基，作为一个亲核基团进攻核酸的磷酸基团，形成 Tyr-磷酸酯中间体连接在暂时断裂的 DNA 一端，在断链的另一端的脱氧核糖上留下 OH。这样 II 型酶可暂时断裂双螺旋 DNA 片段上的一对互补链，产生有酶连接的 DNA 缺口；同时通过拓扑异构酶的结构变化使另一个 DNA 双螺旋片段穿过缺口；第二步与第一步相反，经历了断裂链一端的磷酸基团和另一端的羟基重新连接以及从共价连接的 DNA 上释放酶的步骤，完成一次 DNA 的拓扑转化。

ATP 在旋转酶的催化机制中起到了关键的作用。旋转酶 B 亚基中存在 ATP 酶区域，II 型酶的 ATP 结合和水解、产物释放，驱动着 DNA 双螺旋的链转移过程，这是 II 型酶催化松弛正超螺旋的关键。

拓扑异构酶 II 的 GyrA 亚基能形成 V 型的二聚体，DNA-G 片段(断裂的双链 DNA 片段)通过形成一对 DNA-蛋白质间的磷酸 Tyr 酯键结合于靠近 V 顶端的部位。GyrB 亚基也能形成钳型结构，当其不结合 ATP 时，两个亚基分开，对应酶的开放钳形式。第二个 DNA 双螺旋 T 片段(转移的 DNA 片段)可通过 GyrB 亚基 N 末端的门进入蛋白钳中。结合 ATP 导致酶 GyrB 亚基的两个 N 末端的配对，促使真核 DNA 拓扑异构酶 II 从开放钳到闭合钳的转化。每个酶可结合两个 ATP，结合第一个 ATP 能诱导第二个 ATP 更容易结合。一个 ATP 的结合足够触发酶结构的变化，并关闭 N 末端的蛋白门，从而捕获了 T 片段。然后在酶-G 片段处发生了一系列的结构变化：N 末端的关闭使 GyrA 二聚体与 G-DNA 形成的门打开，T 片段进入 G 片段下的大空腔内。紧接着酶-DNA 复合物关门降低了含有 T 片段的空腔的体积，在蛋白质和 T 片段间产生立体压力，驱动 GyrA 二聚体 C 末端的蛋白门打开，T 片段退出，完成一轮 DNA 双链转移。最后 ATP 水解，释放 ADP、焦磷酸，使 GyrB 亚基 N 末端的门打开，开始新的循环。

无 ATP 时，大肠杆菌旋转酶只能很慢地松弛负超螺旋但不能松弛正超螺旋。但缺少 GyrB N 末端的旋转酶，在无 ATP 的情况下也可松弛正负超螺旋。无 ATP 时，噬菌体 T4 拓扑异构酶表现出低水平的 DNA 转化活性，在高酶浓度时能使超螺旋环打结。而酵母和果蝇的 DNA 拓扑异构酶 II 无 ATP 时，不能进行 DNA 环的拓扑转化。

4. 拓扑异构酶的生物学功能

拓扑异构酶的功能是恢复由一些细胞过程产生的 DNA 超螺旋。在 DNA 活动的许多过程中都能产生正负超螺旋，如在复制过程中，DNA 双螺旋的解开使复制叉的前面产生正超螺旋，在一些条件下，转录能导致聚合酶前产生正超螺旋，在酶后产生负超螺旋。

拓扑异构酶的另一个主要功能是防止细胞的 DNA 过度超螺旋。使细胞内 DNA 的超螺旋程度保持在一个稳定的状态。拓扑异构酶防止 DNA 过度超螺旋不表明细胞中不需要超螺旋的存在。在 DNA 复制起始时需要 DNA 处于负超螺旋状态。生物体内的负超螺旋稳定在 5% 左右，生物体通过多种拓扑异构酶的作用使负超螺旋达到一定的平衡。保持适当的 DNA 超螺旋对 DNA 结构和它与其他分子的相互作用有强烈影响，包括影响原核生物基因的转录起始，保持基因组稳定性，促进重组，抑制有丝分裂。

此外，拓扑异构酶在生物体中还有一些其他功能：例如大肠杆菌中旋转酶有主要的去连环活性；真核拓扑异构酶 II 参与染色体凝缩过程。拓扑异构酶还能解开缠绕的 DNA 双螺旋。

当然，拓扑异构酶也有其不利的方面，如编码大肠杆菌拓扑异构酶 I 的 *topA* 基因突变引起的超负超螺旋可使 DNA 形成 R-环，导致异常的复制。拓扑异构酶也可使 DNA 断裂，影响基因组的稳定性。

（吴东林）

参 考 文 献

[1] Büeler H et al. Cell, 1993. 73:1339～1347
[2] Lindquist S. Cell, 1997. 89:495～498
[3] Horwich A L et al. Cell, 1997. 89:499～510
[4] Farquhar C F et al. Nature, 1998. 391:345～346
[5] Watson J D et al. Nature, 1953. 171:737～738
[6] Watson J D et al. Nature, 1953. 171:964～967
[7] Rich A et al. Ann Rev Biochem, 1984. 53:791～846
[8] Marmur J et al. Prog Nuc Acid Res, 1963. 1:232～300
[9] Courey A J. Cell, 1983. 33:817 -829
[10] Williamson J R. Ann Rev Biophys Biomol Struct, 1994. 23:703～730
[11] Frank-Kamenetskii M D et al. Ann Rev Biochem, 1995. 64:65～95
[12] Wang J C. Ann Rev Biochem, 1996. 65:635～692

第三章　基因与基因组

第一节　基因概念的历史演变[1-2]

基因(gene)是遗传的基本单位。在学习基因的性质之前，首先要区分生物体的表型特征与其潜在的遗传构成之间的差异。生物体的表(现)型(phenotype)是指有机体可见的或可检测的外在性质，而基因型(genotype)是控制这些表型的遗传因子。基因通常是一个稳定的遗传单位，但它也可能发生可遗传的变异，这种变异叫基因突变(mutation)。突变后的基因以新的形式稳定遗传。携带突变基因的生物体叫突变体(mutant)，携带正常基因的生物体叫野生型(wild type)。野生型既可表示基因型也可表示表型。由突变体的表型效应上看，有些突变是无法察觉的，有些突变则可能使其所在的生物体致死。另外，有些遗传性状是由多基因决定的，而另一些性状是由单一基因决定的，有时一个基因也可能产生复杂的表型。

Johannsen 是最早使用基因这个名词的人，可是基因的概念早已隐含在 Mendel 的设想中。Mendel 对豌豆等植物杂交实验进行了研究，认为基因是颗粒状的分散的独立遗传因子，可以从亲代忠实地传递给子代。Mendel 提出了等位基因的概念并总结了两条基本的遗传学定律，即遗传因子的分离定律和自由组合定律。由于当时 Mendel 假定的因子缺乏已知的物理基础而被其工作并未受到重视。遗传的染色体学说提出后，人们意识到染色体在减数分裂和受精时的行为与 Mendel 的遗传粒子的性质类似。由此人们设想，能否把特定的基因定位于特定的染色体上。Morgan 在 1910 年对果蝇突变体的研究为该设想提供了证据。减数分裂时染色体的独立分配解释了位于不同染色体上基因的自由组合现象。但当两个基因位于同一条染色体上时，其行为会偏离 Mendel 的自由组合定律，表现出连锁(linkage)和交换现象。这是因为在一条染色体上有许多基因，它们在染色体上形成了线性的连锁群，相邻的基因可共同遗传。连锁图显示了基因在染色体上的线性排列。至此，Mendel 的基因粒子概念发展为染色体的概念。染色体是一个有许多基因组成的连锁群，基因的线性物理排列是其遗传行为的基础。

1945 年前，基因虽被视为遗传的基本单位，但对基因如何发挥功能并无统一的解释。只能通过检测表型异常的突变体鉴定基因，而突变体的获得是非常偶然的，同时某些突变体不适用于生物化学研究，使基因的研究遇到很大的困难。尽管当时已发现许多生化反应缺陷与特殊突变体相关，但对基因的物理性质及基因突变与生化缺陷之间的关系还不清楚。后来通过对大量突变体代谢途径的研究，1941 年 Beade 和 Tatum 提出了“一个基因一个酶”的假说。它假设每个代谢步骤都是由一个特殊的酶催化的，而这个酶的生成是由一个特定的基因负责的。基因中的突变可导致它所负责的酶活性的改变。直到 1957 年，证明该假说的直接证据才出现。Ingram 通过实验证明，镰状细胞贫血(sickle cell anemia)的病因是编码血红蛋白的基因发生了突变，从而改变了血红蛋白的氨基酸组成。当一种蛋白质是由多亚基构成的聚集体时，需要对这个假说进行一点修正；对于异源多聚体

蛋白质，“一个基因一个酶”假说更精确的表述应该是：“一个基因一条多肽链”。

在遗传学上，也可以用顺反子的概念来表示基因，当两个突变具有相同的表型效应而且其遗传图距很接近时，它们既可能属于一个基因，也可能位于不同的基因上。顺反互补测验（*cis trans* test）（图 3－1）可精确确定两个突变间的关系。其原理是，当两个纯合的亲代突变体杂交时，产生的杂合子后代将遗传两个亲本的突变。如果两个突变位于同一个基因上，杂合子中就不存在野生型的基因，所以具有突变的表型；如果突变位于不同的基因上，在杂合子的每条染色体上相邻的两个基因中，将有一个基因是野生型而另一个基因是突变型，即在杂合子中每个基因都有一个野生型拷贝，因此杂合子表现为野生型的表型，我们把两个基因之间的这种关系称为互补。

杂合子有两种构型，在顺式构型中，两个突变位于同一条染色体上，在反式构型中，两个突变分别位于同源的两条染色体上。上面的互补实验中杂合子为反式构型。两种构型的相对效果取决于两个突变是否位于同一个基因上。当突变位于同一个基因上时，杂合子的表型取决于构型。反式构型中基因在两条同源染色体都有突变，因此表型是突变体。在顺式构型中，一条染色体的基因上有两处突变，而另一条染色体上的基因正常，因此表型是野生型。而当两个突变位于不同的基因上时，表型与杂合子的构型无关。两种情况下，两种基因都有一个野生型的拷贝，表型是野生型。

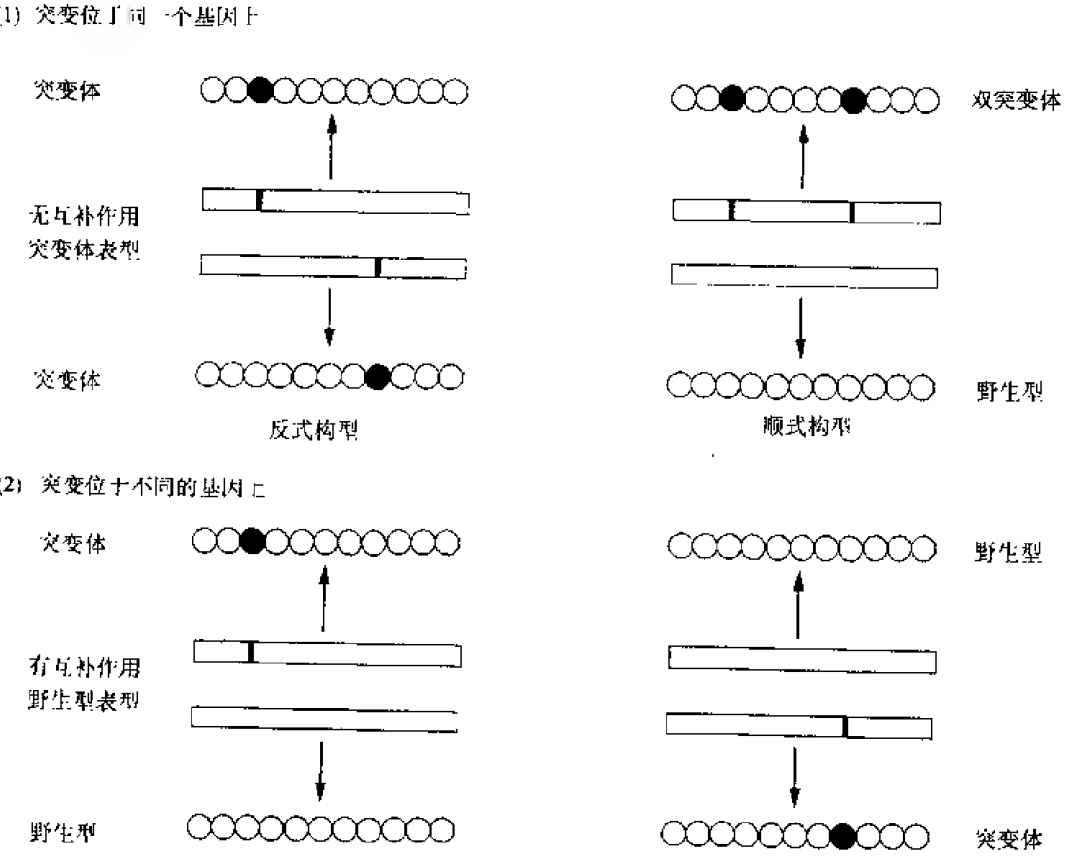


图 3－1　顺反互补测验

当两个突变在反式情况下不能互补时，表明两个突变影响同一生物学功能，因此它们属于一个不可分割的遗传单位，这个单位被称为顺反子(cistron)。发生在一个顺反子中的两个突变在反式构型时不能互补，而互补的发生表明两个突变位于不同的顺反子中。因此，一个顺反子在本质上与一个基因相同，可编码一种多肽链。

第二节　DNA与基因[3]

DNA是遗传的物质基础，基因位于染色体上实际上是位于染色体的DNA上。前面讲到DNA由不同的核苷酸按一定的顺序排列而成，遗传信息就存在于核苷酸序列中。基因被定义为能产生一个特定蛋白质的DNA序列。生物体的表型是通过产生许多特殊的蛋白质来实现的，而蛋白质由一系列氨基酸组成，蛋白质的结构和活性由其氨基酸的排列顺序决定。DNA的核苷酸序列能决定组成它所对应蛋白质的氨基酸序列。DNA序列和相应蛋白质的氨基酸序列之间的关系叫做遗传密码。遗传密码是以三个核苷酸为一组阅读的，每个三核苷酸序列叫做一个密码子(codon)，代表一个氨基酸。基因包括一系列从固定的起始位点读起并在固定的终止位点终止的密码子。通常按5′→3′方向书写，对应于其蛋白质产物的N末端到C末端。在任何给定的区域，DNA双链中只有一条编码蛋白质。通过这种方法，基因就可以使生物体的细胞表现出独特的表型。

现在我们可以理解突变的分子基础。在DNA序列上发生的变化相应地也导致蛋白质氨基酸序列的改变，使蛋白质的生物功能产生变化。DNA上的任何碱基对都可能发生突变，点突变指只有一个碱基对发生变化，导致突变密码子所对应的氨基酸发生改变。DNA上的密码是从一定的起点按特定的可读框阅读的，插入或缺失一个核苷酸都会造成移位突变使随后的可读框发生改变，导致在突变位点后的氨基酸序列全部发生变化。一些DNA位点上更容易发生突变，这些位点叫做突变热点(mutation hotspot)。突变的存在更加确定地表明DNA是遗传物质。同时，如果一个基因突变能导致一个蛋白质的变化，我们就可以推论该基因编码这个蛋白质。

通过比较基因的核苷酸序列和相应蛋白质的氨基酸序列，我们可以确定在细菌和病毒中基因的核苷酸序列与蛋白质产物中的氨基酸序列是精确对应的。每个基因都包含连续的DNA串，其长度与所表达蛋白质的氨基酸数目直接相关。根据遗传密码，编码一个有n个氨基酸的蛋白质需要一个有$3n$个碱基对的基因。

在比较基因和蛋白质序列的相互关系时，我们只考虑了直接对应于蛋白质的DNA序列，但是基因并不能直接翻译成蛋白质，而是通过产生一个信使RNA中间体来进行蛋白质的合成。信使RNA的序列与DNA一条链的序列互补而与另一条序列相同(除了用U取代T)。信使RNA中包含对应于蛋白质中氨基酸序列的一段核苷酸序列。这部分核酸序列叫做编码区(coding region)。但在信使RNA的每个末端还有额外的序列，这些序列并不直接代表蛋白质。有时在这些额外的非编码区中发生的突变也能阻碍基因的功能，表明这些区域属于遗传单位的合理部分(图3-2)。所以可以认为基因对应于信使RNA中的整个序列，它既包括编码蛋白质的序列也包括编码区两侧的序列。结构基因被定义为一段用来产生特定蛋白质的连续DNA序列，对应于成熟mRNA 5′到3′末端之间的区域。

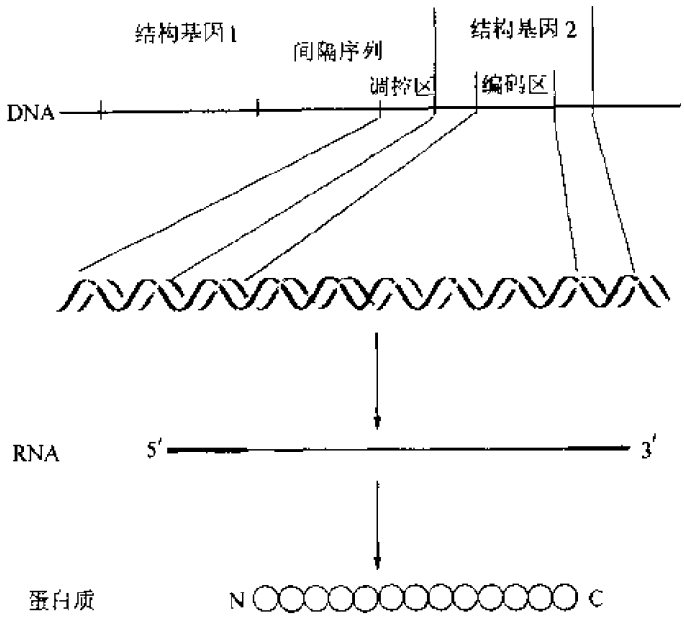


图 3-2 DNA 与基因的关系

DNA 分子上不同的区域有不同的功能，并不是所有的序列都具有编码功能，在原核生物中结构基因占整个 DNA 的大部分，而在真核生物中可能只占一小部分。除了上述基因序列外，在结构基因间还含有一些没有编码功能的间隔区(spacer region)，其中包括一些复制、转录、翻译过程的调控信号，即可被调控分子识别的序列。一个基因是否表达是受与邻近编码区域的 DNA 序列结合的蛋白质控制的，只有蛋白质结合 DNA 时，信使 RNA 才能合成；当与蛋白质结合的 DNA 序列上发生突变时，蛋白质不能识别 DNA，结果 DNA 不能表达。这里 DNA 的功能是由它的序列直接决定的，而不是通过编码间接决定的。所有这些区域，包括表达蛋白质的基因和上述的调控序列都构成了遗传信息。因此我们可以把遗传单位的概念从基因扩展为转录单位，转录单位(transcription unit)包括转录的启动子及其上游的其他调控区域、基因本身和转录的终止序列。

除了编码蛋白质的基因，还有最终产物是 RNA 的基因，如编码 tRNA 和 rRNA 的基因，这些基因的产物与蛋白质的生物合成有关。所以基因是具有一定生物功能的 DNA 序列，基因主要位于染色体上，除了染色体 DNA 外，细菌的质粒、真核生物的叶绿体、线粒体等细胞器都含有 DNA，其上也含有基因，这些染色体外的 DNA 称为染色体外遗传物质。

通常情况下 DNA 的结构是十分稳定的，基因位于染色体上的特定部位。但有些 DNA 序列不但可以在染色体上移动，而且还可以从一个染色体跳到另一个染色体，甚至可以从一个细胞跳到另一个细胞。在这些 DNA 序列跳跃的过程中往往导致 DNA 链的断裂和重接，这些可以自由移动的 DNA 序列称为转座子。

至此我们对基因与基因组 DNA 的关系有了初步的认识，但是要准确了解其实际情况，我们还要确定许多问题，例如基因两侧有多少序列与其功能有关；相邻的两个基因距

离多远；在基因组中有多少成分是转录单位；是否相邻的基因是功能上相关的；有无一种调控机制可以调控染色体上比单个基因更长的区域

第三节　真核生物的割裂基因[4]

一、割裂基因的发现

遗传密码信使 RNA 总是对应于编码其蛋白质产物的核苷酸序列，但真核生物基因在编码区域内往往含有额外的序列，打断了对应于蛋白质的序列。1977 年人们首先发现真核生物的基因是不连续的，大大改变了原来对基因结构的看法，现在知道大多数真核生物的基因都是不连续基因或割裂基因(split gene)。所谓割裂基因是指基因的编码序列在 DNA 分子上不是连续排列的，而是被不编码的序列所隔开。构成割裂基因的 DNA 序列被分为两类：基因中编码的序列称为外显子(exon)，外显子是基因中对应于信使 RNA 序列的区域；不编码的间隔序列称为内含子(intron)，内含子是从信使 RNA 中消失的区域。割裂基因由一系列交替存在的外显子和内含子构成，基因两端起始和结束于外显子，对应于其转录产物 RNA 的 5′和 3′末端。如果一个基因具有 n 个内含子，则相应地含有 $n+1$ 个外显子。

存在割裂基因的最初证据是通过比较 DNA 和相应的信使 RNA 结构获得的。割裂基因可通过 mRNA 与 DNA 的分子杂交而发现，杂交后用电镜进行观察，如果基因中含有内含子，因为其 mRNA 中没有相应的序列，在所形成的 RNA-DNA 杂交双链的某一部位就会出现不能配对的单链环，这就是内含子。用这个方法还可以确定内含子在基因中的位置和大小(图 3-3)。

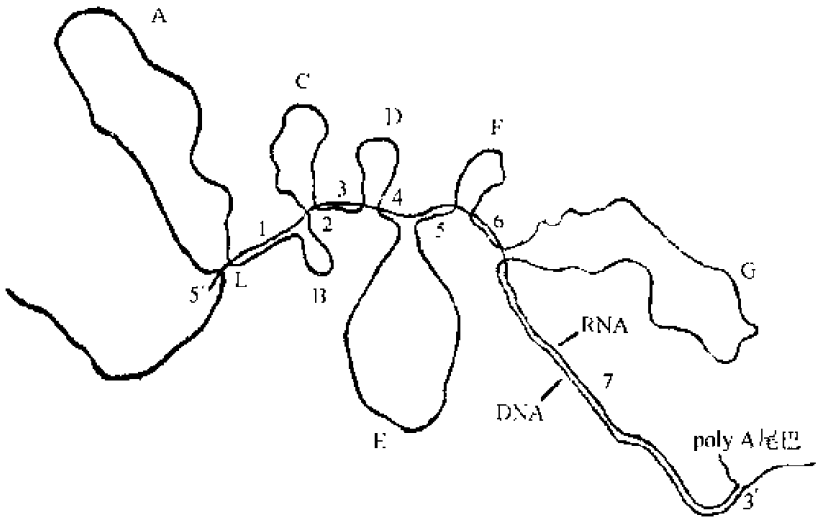


图 3-3　割裂基因

真核生物基因表达的过程是一个不同于原核生物的过程，与原核生物相比多了一个新的步骤。DNA 经过转录产生了精确对应于基因组 DNA 序列的 RNA 拷贝，但这个 RNA 只是一个前体，不能直接用于表达蛋白质。在成熟的 mRNA 中只由外显子存在，所

以必须从 RNA 原初转录产物中去除内含子序列以产生一个只由外显子构成的信使 RNA。从原初转录产物中除去内含子的过程叫做 RNA 的剪接(splicing)(图 3-4),首先从原初转录产物上精确地切除内含子,然后把切口两边对应于外显子的 RNA 序列的末端共价连接形成完整的分子。经过剪接后,所有的外显子按其在 DNA 上相同的顺序连接在同一个 RNA 分子上。剪接是一个分子内的反应,来源于不同 RNA 分子的外显子不能连接到一起。

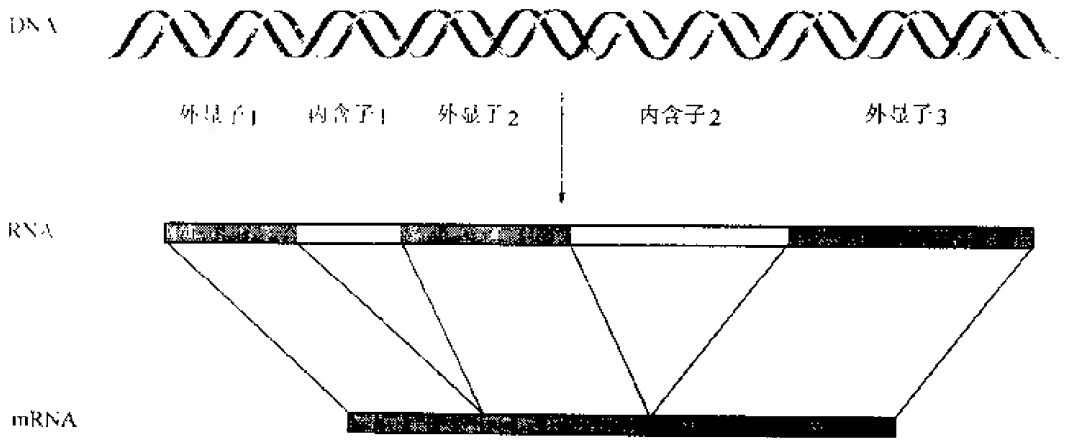


图 3-4 割裂基因的剪接

二、割裂基因的分布

基因的断裂现象广泛地存在于各种生物不同的基因中:包括编码蛋白质的核基因、编码 rRNA 的核基因,以及编码 tRNA 的核基因,在某些低等真核生物的线粒体以及叶绿体基因中也发现了断裂现象。基因的不连续现象不仅存在于真核生物基因组中,甚至在某些原核生物如古细菌和大肠杆菌的噬菌体中也有发现。这种 DNA 和相应信使 RNA 结构上的差异在真核生物中是普遍的,但在细菌中却十分少见。在真细菌的基因组中不存在割裂基因,真核生物的基因不一定都中断,也有一些结构基因不含内含子,与原核生物的基因一样直接对应于其蛋白质产物,如组蛋白基因和干扰素基因等,在酵母中大部分基因事实上是不中断的。在高等真核生物中,大多数基因是中断的。

三、割裂基因的性质

割裂基因的内含子无论在数量和大小上都有很大不同,但大多数割裂基因都有共同的性质:割裂基因的外显子在基因中的排列顺序和它在成熟 mRNA 产物中的排列顺序是相同的;某种割裂基因在所有组织中都具有相同的内含子成分;核基因的内含子通常在所有的可读框中都含有无义密码子(nonsense codon),因此一般没有编码功能。内含子上发生的突变不能影响蛋白质的结构,所以其突变往往对生物体是没有影响的;但也有例外,

例如一些发生在内含子上的突变可通过抑制外显子的相互剪接阻止信使 RNA 的产生。

利用结构基因的特殊 DNA 限制片段作为探针，我们可以检测基因组中与之有亲缘关系的序列，结果表明一个基因的外显子常与其他基因的外显子有亲缘关系。如果一个外显子可与其他基因的外显子片段互补，则表明这两个基因可能起源于共同祖先，经过基因扩增后，在进化过程中逐渐扩大差距而形成不同的基因。两个相关基因内含子之间的亲缘关系远远不如其外显子之间的亲缘关系紧密，有些含有相同外显子的基因其内含子的差异是如此之大，以致于不能找到序列间的亲缘关系。这是因为在进化过程中，相关基因的内含子比外显子变化得更快。其实，突变是以相同的频率发生在外显子和内含子上的，但是发生在外显子上的突变将使基因编码产物失去功能，导致生物体无法生存，经过长期选择这种突变就从外显子中除去了；而内含子由于没有来自编码功能的限制，可以自由地积累各种突变，导致产生较大的变化。

第四节　基因大小

由于割裂基因的存在，人们认识到基因比实际编码蛋白质的序列要大得多。外显子的大小与基因的大小没有必然的联系。与整个基因相比，编码蛋白质的外显子要小得多，大多数外显子编码的氨基酸数小于 100。不同种类的生物体中外显子的大小并没有明显的不同，基因可能是由一些小的、编码较小的独立蛋白质分区的单位在进化过程中加合起来的。也有一些比较大的外显子编码不翻译的 5′和 3′区域。

许多长基因的存在并不能表明其编码产物较长，而是因为其含有较长的内含子。比较基因的长度和由它们编码的 mRNA 的长度，我们可以看出在酵母中信使分子大小的变化范围与其基因大小的变化范围并没有什么不同。但是在哺乳动物中情况有明显的不同，它们的信使分子常常不到 10 kb，而与之相比基因的大小常常达到 100 kb。

基因的大小取决于它所包含的内含子的长度，一些基因的内含子特别长，例如哺乳动物的二氢叶酸还原酶基因含有 6 个外显子，其 mRNA 的长度为 2000 bp，但它含有极长的内含子，使其基因的总长度长达 25～31 kb。内含子通常比外显子大很多，导致整个基因比其编码区域大很多。而且内含子之间有很大不同，它们的大小从 200 个碱基对左右到上万个碱基对。在一些极端的例子里，甚至有 50～60 kb 的内含子。

通常在进化上相关的基因有相似的组织形式，甚至内含子在基因中的位置也是保守的。它们之间基因长度的变化主要取决于内含子的长度的变化。在所有已知的珠蛋白基因中，断裂都发生在相同的位置，第一个内含子位于 30 和 31 位氨基酸密码子之间，第二个内含子位于 104 和 105 位氨基酸密码子之间。通常第一个内含子总是很短，第二个内含子要长得多，但不同物种之间具体的长度有所不同。而外显子的长度却基本一致，不同珠蛋白基因其大小的差异是因为第二个内含子长度的差异造成的。例如小鼠的 α-珠蛋白基因第二个内含子只有 150 bp 长，所以 α-珠蛋白基因的总长度为 850 bp，而 β-珠蛋白基因的长度为 1382 bp，但二者的 mRNA 长度却相差不多。（α-珠蛋白的 mRNA 为 585 碱基，而 β-珠蛋白 mRNA 为 620 碱基）。基因的大小还与它所包含内含子的数目有关。在不同的基因中，内含子的数目变化很大，有些割裂基因含有一个或少数几个内含子，如珠蛋白基因；某些基因含有较多的内含子，例如鸡卵清蛋白基因含有 7 个内含子，伴

清蛋白基因含有 16 个内含子。

在进化过程中，割裂基因在较低等的真核生物中开始出现。在低等真核生物酿酒酵母中，大多数基因是非断裂的，而且割裂基因含有外显子的数目也是非常少。实际上很少有酵母的基因超过 4 个外显子，而且它们的长度都很短。真菌基因的外显子也较少，小于 6 个，而且它们的长度也很短，不到 5 kb；在更高等的真核生物中，如昆虫和哺乳动物中，情况刚好相反：只有极少量的基因含有非断裂的编码序列。昆虫的外显子比较少，一般不超过 10 个。哺乳动物则比较多，一些种类甚至有几十个外显子。

由于基因的大小取决于内含子的长度和数目，导致酵母和高等真核生物的基因大小有很大的不同。大多数酵母基因小于 2 kb，很少有超过 5 kb 的。与此相反，在高等的真核生物中，开始出现长的基因，蝇类和哺乳动物基因很少小于 2 kb，大多数长度在5～100 kb 之间。但当基因的长度大到一定程度后，DNA 的复杂性与生物体的复杂性之间开始失去必然的联系。例如虽然属于同一个门，果蝇细胞的 DNA 总量较小而家蝇细胞的 DNA 总量却是它的 6 倍。而且在较高等的真核生物中，基因大小与外显子的大小和数目之间也没有必然联系。基因的大小主要取决于它所包含的内含子的长度。

表 3－1 总结了一系列生物体的平均基因大小，当比较不同物种时，我们可以看出从低等真核生物到高等真核生物的 mRNA 和其基因的平均大小略有增加，平均外显子数目的明显增加是真核生物的一种标志。在哺乳动物、昆虫、鸟类中，基因的平均长度将近是其 mRNA 长度的 5 倍。

表 3－1　不同生物的平均基因大小

种类	平均外显子数目	平均基因长度(kb)	平均 mRNA 长度(kb)
酵母	1	1.6	1.6
真菌	3	1.5	1.5
藻虫	4	4.0	3.0
果蝇	4	11.3	2.7
鸡	9	13.9	2.4
哺乳动物	7	16.6	2.2

第五节　重叠基因[5]

在某些情况下，基因与蛋白质产物间的关系并不是一成不变的，从同一段 DNA 序列上可得到不同的 mRNA 产物以编码多种蛋白质。

一、原核生物的重叠基因

在细胞基因中，一般一段 DNA 序列只以三种蛋白质可读框的一种被阅读，但是在一些病毒或线粒体基因中，两个邻近的基因以一种巧妙的方式发生重叠，并以不同的可读框被阅读并表达，因此一段相同的 DNA 序列可以编码两个非同源蛋白质。重叠的距离通常相对较短，所以大部分序列对应的蛋白质仍然有独特的编码功能。$\Phi X174$ 的 DNA 序列组织上最显著的特点是有重叠基因(overlapping gene)和基因内基因(图 3－5)。基因 *B*

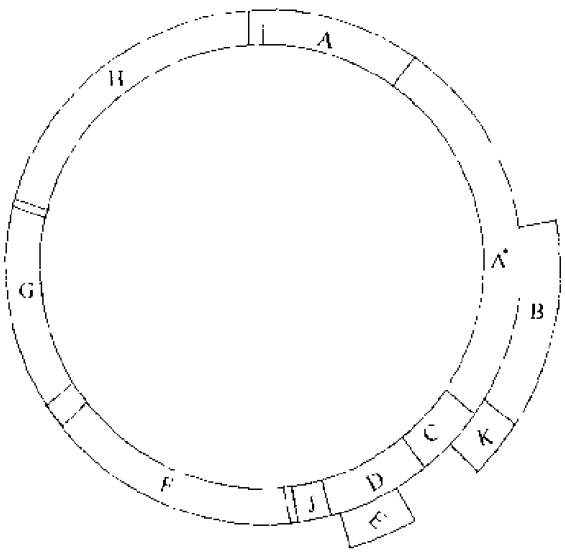


图 3-5 ΦX174 的重叠基因

和 K 重叠在基因 A 之内，基因 E 重叠在基因 D 之内。这些重叠在一起的基因在表达时使用不同的可读框，因此虽然是同样的 DNA 序列，但表达出来的是不同的蛋白质。在重叠基因的情况下，在共同序列上发生的突变可能影响其中一个基因的功能，也可能影响两个基因，所以它可能属于一个互补群也可能属于两个互补群。

重叠基因还有一种相对简单的情况：一个基因包含在另一个基因之中，两个基因使用相同的可读框。小基因可独立地表达一种蛋白质，相当于整个基因表达的蛋白质的一部分。这种情况下，不存在特殊的遗传问题，最终的结果类似于一个完整的蛋白质发生了部分的断裂。

重叠基因及基因内基因的现象反映了原核生物利用有限的遗传物质表达更多生物功能的能力

二、真核生物的重叠基因

通常割裂基因的每个外显子编码一段单一的氨基酸序列，对应于整个蛋白质上的相应部分，而内含子不在最终的蛋白质产物中表达，二者的作用是迥然不同的。但是有些基因中内含子和外显子的定义是相对的，与它表达的途径有关。

在这些基因中，选择性的基因表达模式引起了外显子连接途径的转变。所以一个特定的外显子可能选择性地与不同的外显子连接形成信使 RNA。这种选择性形式产生的两种蛋白质中，一部分相同而其他部分不同。一段区域以一种途径表达时作为外显子，而以另一种途径表达时作为内含子。因为此时一段 DNA 序列通常以多种方式起作用，所以不能被简单地称为外显子或内含子。

如图 3-6 所示，由于选择性剪接使得外显子 4 序列存在于一个 mRNA 中，而在另一个 mRNA 中不存在。在第一种情况下，它被定义为外显子；然而在第二种 mRNA 的剪接时，它却被作为外显子 3 和 5 之间内含子的一部分而被切除。这种选择性剪接可从一段 DNA 序列产生有部分重叠序列的多种蛋白质。

如大鼠的肌钙蛋白 T 基因就是利用这种方式产生 α、β 两种产物。在其表达时分别使用了 α、β 两种外显子。真核生物采用这种方式是很奇怪的，因为一方面高等真核生物的基因组十分巨大，无法用已知的功能解释如此之多的 DNA 含量。但同时它却用同一段 DNA 序列产生多种产物。

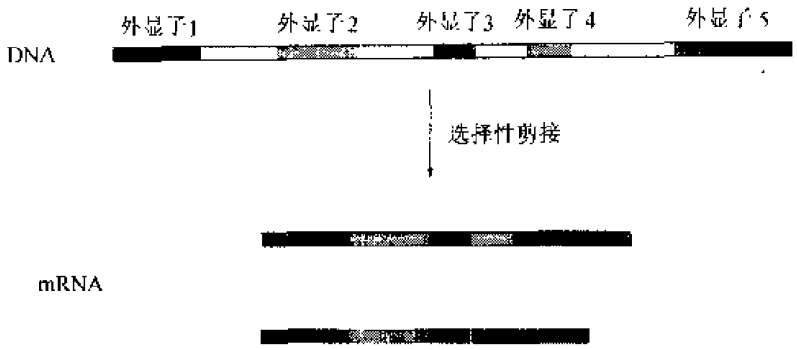


图 3-6 选择性剪接可产生不同的 RNA 产物

第六节 真核生物的基因组[6]

一、真核生物的基因组

一个物种的单倍体的染色体数目称为该物种的基因组(genome)。与原核生物相比，真核生物基因组有一些独特的性质。真核生物基因组的主要成分被核膜包围在核内，与细胞质分开。组成核基因组的序列形式也与原核生物不同，不同物种核 DNA 的数量差异很大，染色体数目也不同，而且细胞器基因组的大小也有很大差异。

另外在真核生物中存在割裂基因，而且有些序列有许多相同的拷贝。在真核生物中还有大量不编码蛋白质的 DNA 序列。由于被分成核区和细胞质，真核生物的基因表达形式也必然与原核生物不同。

表 3-2 人的基因组组成

项目	部位	大小
基因数目	核	65 000 - 80 000
	线粒体	37
基因密度	核	40～45 kb
	线粒体	0.45kb
平均基因大小		10～15 kb
基因间距离		25～30 kb
平均外显子大小		170 bp
mRNA 大小	5′非翻译序列	0.1 kb
	编码区	1.5 kb
	3′非翻译序列	0.6 kb

二、基因组大小与 C 值矛盾

一个单倍体基因组的全部 DNA 含量总是恒定的。这是物种的一个特征，通常称为该物种的 C 值。不同物种的 C 值差异很大，最小的枝原体只有 10^6 bp，而最大的如某些显花植物和两栖动物可达 10^{11} bp，图 3-7 为不同生物种类的 C 值范围分布图。由图可见，随着生物的进化，生物体的结构和功能越来越复杂，其 C 值就越大，例如真菌和高等植物同属于真核生物，而后者的 C 值就大得多。这一点是不难理解的，因为结构和功能越复杂，需要的基因产物的种类越多，也就是说需要的基因越多，因而 C 值越大。

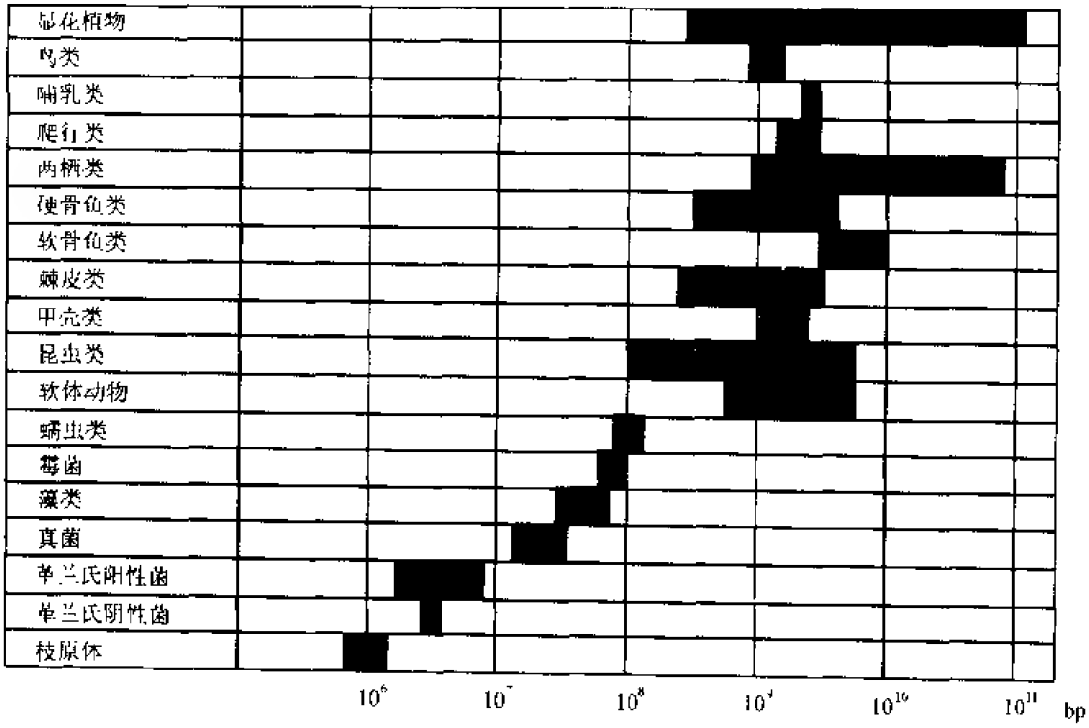


图 3-7　不同种类生物基因组的 C 值分布

然而另一方面，随着进一步的进化，生物体复杂性和 DNA 含量之间的关系变模糊了，出现了很多令人不解的现象。一些物种基因组大小的变化范围很窄。鸟、爬行动物、哺乳动物各门内基因组大小的范围只有两倍的变化。但大多数昆虫、两栖动物和植物的情况却不同，在结构、功能很相似的同一类生物中，甚至在亲缘关系十分接近的物种之间，C 值可以相差数十倍乃至上百倍。突出的例子是两栖动物，C 值小的可以低至 10^9 bp 以下，C 值大的可以高达 10^{11} bp。而哺乳动物的 C 值均为 10^9 bp 的数量级。人们很难相信为了区别不同的两栖动物，所需基因的数量会有 100 倍的差别，而且两栖动物的结构和功能会比哺乳动物更复杂。

由于人们无法用已知功能来解释基因组的 DNA 含量，所以产生了 C 值矛盾（C value paradox，又称 C 值悖理）。它表现在两个方面：一个方面是，与预期的编码蛋白质的基因的数量相比，基因组 DNA 的含量过多。另一个方面是一些物种之间的复杂性变化范围

并不大，但是 C 值却有很大的变化范围。这些问题的解决有待于进一步的研究。

三、基因组的基因数目

从已知基因的数据，我们可以知道基因的平均大小。如果知道基因组总体大小，我们就能粗略地计算出基因组中基因的数目。虽然一些基因通过选择性的表达可产生一个以上的产物，但这种现象并不常见，对基因组数目的计算影响不大。

由于 DNA 中存在非编码序列，使计算产生误差，故须确定基因密度。为准确地确定基因数目，需要知道整个基因组的 DNA 序列。目前已知酵母基因组的全序列，其基因密度较高，平均每个可读框（open reading frame，ORF）为 1.4 kb，而基因间的平均分隔为 600 bp，即大约 70% 的区域由可读框占据。其中约一半基因是已知的基因或与已知基因有关的基因，其余是新基因，因此可推测未发现基因的数目。由此推知细菌可能有 2400 个基因，酵母为 6000 个，果蝇为 9000 个，哺乳动物高达 125 000 个。

表 3-3　不同生物的基因数目

种类	基因组大小(bp)	基因数目
枝原体	1.0×10^6	750
噬菌体 T4	1.6×10^5	200
大肠杆菌	4.2×10^6	2350
酵母	1.3×10^7	6100
果蝇	1.4×10^8	8750
人	3.3×10^9	65 000～80 000

如果不知道基因组的基因密度，就难以估计整体数目。人们通过基因分离鉴定知道一些物种的基因数目，但这只是一个最小值，真正的基因数目往往大得多。此外通过测序鉴定可读框也可以推测基因数目，但这种方法也有缺陷：例如有的可读框可能不是基因，尤其当可读框很短时；而且一些基因的外显子在分离时可能会断裂，这都导致过高估计基因数目。所以鉴定所有的可读框可得到基因数目的最大值。

另一种测定基因数目的方法是计算表达基因的数目。在脊椎动物细胞中平均表达 10 000～20 000 个基因。但由于在细胞中表达的基因只占有机体整个基因数目的一小部分，所以这个方法也不能准确估计基因数目。一般真核生物的基因是独立转录的，每个基因都产生一个单顺反子的信使。秀丽隐杆线虫（*C. elegans*）的基因组是个例外，其中 25% 的基因能产生多顺反子的转录产物，通过剪接表达出多种蛋白质。这种情况会影响对基因数目的测定。

四、必需基因和基因总数

另一个与基因数目有关的工作是通过突变分析确定必需基因的数量。如果在染色体一段区域充满致死突变，通过确定致死位点的数量就可得知这段染色体上必需基因的数

量。然后外推至整个基因组，我们可以计算出必需基因的总数。利用这个方法，计算出果蝇的致死基因数量为5000。如果果蝇和人的基因组的情况相同，我们有可能预测人有100 000个以上致死基因。但测定的致死位点，即必需基因的数目必然小于基因总数。目前我们还无法知道非必需基因的数量，通常基因组的基因总数可能与必需基因的数量处于相同的数量级。

通过在酵母中确定必需基因比例的实验发现：当在基因组中随机引入插入突变时，只有12%是致死的，另外的14%阻碍生长，大多数插入没有作用。因为插入序列携带了转录终止信号，因此应该阻碍所插入基因的表达。所以酵母表达基因中40%是非必需基因。这是因为存在着冗余基因，即许多基因是多拷贝的。但我们无法确定多少基因是唯一的，多少基因是冗余的。

第七节　真核生物DNA序列组织[7~13]

一、真核生物DNA的复性动力学

复性过程是一个复杂的多步反应过程。复性的第一步是当两条单链相互碰撞时，假如在某个区域遇到正确的配对，就会先形成一个双链核心。核心形成后互补链能很快地相互配对，完成复性过程，核心的形成是限速步骤。复性过程基本上符合二级反应动力学，故称复性动力学(reassociation kinetics)。单链消失的速度可用下面的公式表示：

$$\frac{-dC}{dt} = kC^2$$

其中C为t时间时单链DNA的浓度，k为二级反应常数。上面的公式可以重排为：

$$\frac{-dC}{C^2} = k\,dt$$

对上式积分得

$$\frac{1}{C} - \frac{1}{C_0} = kt$$

可重排为

$$\frac{C}{C_0} = \frac{1}{1 + kC_0t}$$

此公式表明反应中单链DNA的分数$\frac{C}{C_0}$是C_0t，即初始浓度和反应时间乘积的函数(图3　8)。

在一个特定的实验中，C_0是已知的，C的浓度是可以测定的，如以$\frac{C}{C_0}$对C_0t作图，可得到如图3-8所示的曲线，称为C_0t曲线。当复性进行一半时，$\frac{C}{C_0} = 1/2$，这时的C_0t值定义为$C_0t_{1/2}$，可以看出$C_0t_{1/2} = \frac{1}{k}$。如前所述，很多因素会影响复性的速度，但我们如果把所有的可变因素包括温度、离子强度、片段的大小都控制起来，那么剩下的唯一可变因素就是DNA序列的复杂性(complexity)X。DNA分子序列的复杂性定义为：最长没有重复序列的核苷酸对数值。复杂性与$C_0t_{1/2}$成正比，$X = kC_0t_{1/2}$。对于在DNA中不含重复序列的有机体来说，X就是以核苷酸对所表示的基因组的大小。两种不同复杂性的DNA

序列，当DNA的绝对含量（即总的核苷酸数）相同时，则复杂性小的DNA分子浓度高，复性就快，需要的时间 t 就短，因而 $C_0t_{1/2}$ 值就小；而复杂性大的DNA复性慢，$C_0t_{1/2}$ 值就大。以大肠杆菌的 $C_0t_{1/2}$ 为标准，就可以计算出待测样品的复杂性，其计算公式为：

$$X=\frac{4.2\times10^6\ \text{bp}\times\text{样品 DNA}C_0t_{1/2}}{\text{大肠杆菌 DNA}C_0t_{1/2}}。$$

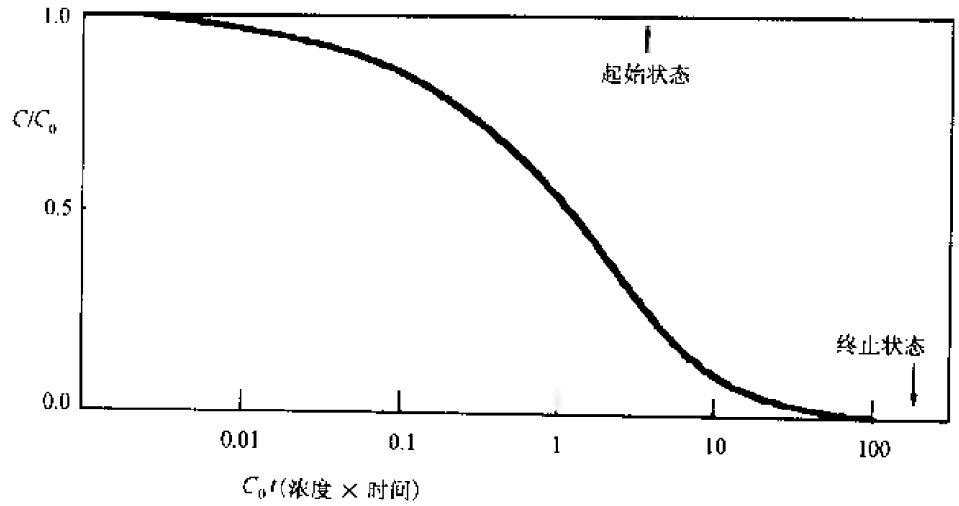


图 3-8 Cot曲线

从图3-9可以看出，各种原核生物DNA的Cot曲线形状都很相似，跨度一般只有两个数量级，只是 $C_0t_{1/2}$ 值不同。表明各种原核生物的DNA都是单一序列，只是其序列复杂性不同。

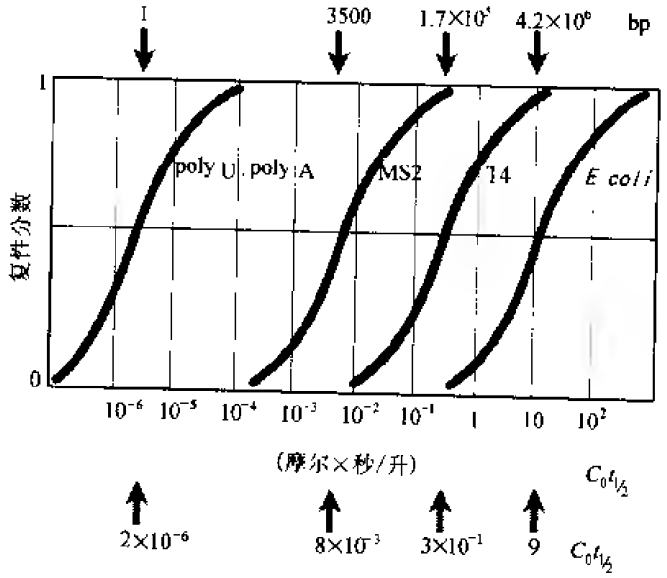


图 3-9 不同种类原核生物的Cot曲线

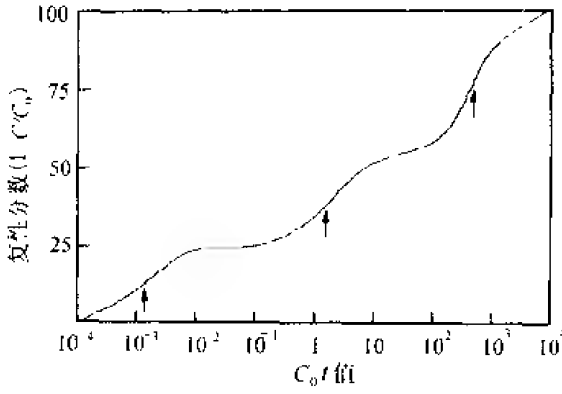


图 3-10 真核生物 DNA 复性动力学模式图

真核生物的 DNA 复性情况就不同了，不但曲线的形状不同，而且常常跨越 7～8 个数量级。图 3-10 是一个真核生物 DNA 复性曲线的模式图。可以看出，复性反应分三个组分进行，每个组分代表基因组中不同复杂性的序列类型，整个基因组 DNA 为 7.0×10^8 bp。第一相为快复性组分，代表了 25% 的总 DNA，C_0t 值范围为 $10^{-4}\sim2\times10^{-2}$，$C_0t_{1/2}=0.0013$。第二相为中间复性组分，代表了 30% 的总 DNA，C_0t 值范围为 0.2～100，$C_0t_{1/2}=1.9$。第三相为慢复性组分，占总 DNA 的 45%，C_0t 值范围为 80～10 000，$C_0t_{1/2}=630$。

由于体系中存在多种组分，每一组分的实际浓度与单独复性时不同。如对于第三组分来说，由于它占总 DNA 的 45%，其浓度 C_0 也是总 C_0 的 45%，因而它实际的 $C_0t_{1/2}$ 应该为 $630\times45\%=283$，就是说如果把这一组分提取出来单独作复性实验，其 $C_0t_{1/2}=283$。假定在同样条件下大肠杆菌的 $C_0t_{1/2}=4.0$，就可以计算出第三组分的复杂性为 3.0×10^8 bp，同样可以算出第二组分的复杂性为 6×10^5 bp，第一组分为 340 bp。

其中每一组分的重复频率 $f=\dfrac{\text{基因组 DNA 总长度}\times\text{这个组分占的比例}}{\text{复杂性}}$。由此可算出第三组分的重复频率 f 为 $\dfrac{7.0\times10^8\ \text{bp}\times45\%}{3.0\times10^8\ \text{bp}}=1$。而第二组分 f 为 350，第一组分 f 为 50 000。

在实际工作中，往往不知道基因组的大小，这时也可利用以上的公式，假设复性最慢的组分为单一序列，重复频率 $f=1$，从而计算出基因组大小的近似值。

根据 DNA 复性动力学的研究，真核生物的 DNA 序列可以分为 4 种类型：

(1)单拷贝序列　又称非重复(nonreplicative)序列，在一个基因组中只有一个拷贝，真核生物的大多数基因都是单拷贝的。

(2)轻度重复序列　在一个基因组中有 2～10 个拷贝(但 2～3 个拷贝常常被视为非重复序列)，如组蛋白基因和酵母 tRNA 基因。以上二者对应于慢复性组分。

(3)中度重复(moderately repetitive)序列　对应于中间复性组分，有十至几百个拷贝。中度重复序列一般是不编码的序列，它们可能在基因调控中起重要的作用，包括开启或关闭基因的活性、促进或终止转录、DNA 复制的起始等。这些重复序列的平均长度大约 300bp。它们在一起构成了序列家族。与非重复序列相间排列。

(4)高度重复(highly repetitive)序列　对应于快复性组分，有几百个到几百万个拷贝。其中，有一些是重复数百次的基因，如 rRNA 基因和某些 tRNA 基因；大多数是重复程度更高的序列。

不同生物中非重复基因占基因组的比例差别很大。原核生物含有完全不重复的 DNA，低等真核生物的大部分 DNA 是非重复的。重复组分不超过 20%，而且基本是中等程度重复组分。在动物细胞中，接近一半的基因组 DNA 被中等或高度重复的组分占据。

在植物和两栖动物中，非重复的 DNA 只占基因组的很小一部分，中等和高度重复的组分高达 80%。

真核基因组的序列组织形式千差万别。在一些多倍体植物中没有非重复的 DNA，复性最慢的组分也有 2～3 或更多的拷贝。而在螃蟹的基因组中，没有中等重复的 DNA，只有高度重复和非重复的 DNA。在低等真核生物中，没有高度重复的 DNA。

在简单的生物体中，非重复 DNA 的长度随基因组大小的增加而增加。但很少有非重复 DNA 组分超过 2×10^9 bp 的生物体。当基因组的大小在 3×10^9 bp(如哺乳动物)以上时，随着基因组大小的进一步增加，非重复 DNA 组分将不再增加，只是重复组分的数量和比例的增加。基因组非重复 DNA 的含量与有机体的复杂性一致。如大肠杆菌为 4.2×10^6 bp，秀丽隐杆线虫为 6.6×10^7 bp，果蝇约为 1×10^8 bp，哺乳动物为 2×10^9 bp。

二、真核生物的单一序列

利用 RNA 作探针进行的 RNA-DNA 之间的核酸杂交可鉴定基因组中的结构基因。在 DNA 驱动的杂交反应中(DNA 过量)，只有一小部分 RNA(小于 10%)的 $C_0t_{1/2}$ 对应于中等重复序列，而大部分组分的 $C_0t_{1/2}$ 对应于非重复的 DNA。通常占总 RNA 量的 50%。那些没有测出杂交反应的组分可能也代表非重复序列。也就是说，绝大多数 mRNA，可能高达 80%，是同非重复 DNA 组分结合的。表明大多数结构基因位于非重复 DNA 序列上，即基因组中的非重复 DNA 序列决定生物体的复杂性。

三、真核生物的重复序列

基因组的很大一部分是由一系列紧密相关的非等位 DNA 序列构成，称为 DNA 序列家族 DNA sequence family 或重复 DNA。其中包括有编码功能的基因家族，也包括没有编码功能的重复 DNA 序列家族。

(一) 基 因 家 族

1. 基因家族和基因簇

基因家族(gene family)是真核生物基因组中来源相同，结构相似，功能相关的一组基因。尽管基因家族各成员序列上具有相关性。但它们序列相似的程度以及组织方式不同。其中大部分有功能的家族成员之间相似程度很高。但也有些家族成员间的差异很大，甚至还有无功能的假基因(pseudogene)。基因家族的成员在染色体上的分布形式是不同的，其中一些基因家族的成员在特殊的染色体区域上成簇存在，而另一些基因家族的成员在整个染色体上广泛地分布，甚至可存在于不同的染色体上。根据家族成员的分布形式可把不同的基因家族分为成簇存在的基因家族(clustered gene family)或基因簇以及散布的基因家族(interspersed gene family)。

1) 基因簇(gene cluster)：基因家族的各成员紧密成簇排列成大段的串联重复单位，

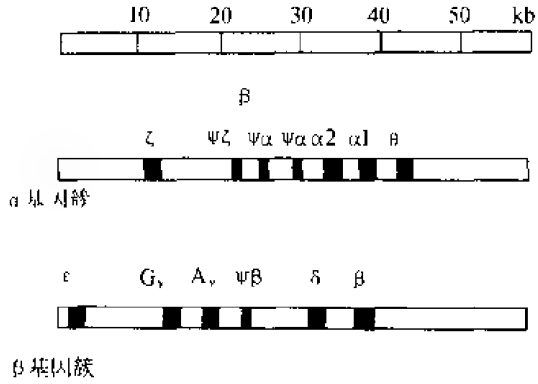


图 3-11 人类类 α 珠蛋白和类 β 珠蛋白基因家族分别组成基因簇，每个基因簇都包含有功能的基因和假基因。

定位于染色体的特殊区域。它们是同一个祖先基因扩增的产物。也有一些基因家族的成员在染色体上的排列并不十分紧密，中间可能包含一些无关的序列。但总体来说，它们分布在染色体上相对集中的区域。基因簇中也包括没有生物功能的假基因。通常基因簇内各序列间的同源性大于基因簇间的序列同源性（图3-11）。

2）散布的基因家族：家族成员在 DNA 上无明显的物理联系，甚至分散在多条染色体上。各成员在序列上有明显差别，其中也含有假基因。但这种假基因与基因簇中的假基因不同，它们来源于 RNA 介导的转座。

2. 广义的基因家族

按照基因家族成员之间序列相似的程度，广义上基因家族可分为以下几类：

1）一类是经典的基因家族，家族中各基因的全序列或至少编码序列具有高度的序列同源性。例如 rRNA 基因家族和组蛋白基因家族。基因家族的各成员之间有极高的序列同源性。这是因为在进化过程中，家族成员有自动均一化的趋势。它们的特点是：①各成员间有高度的序列一致性，甚至完全相同。②拷贝数高，常有几十个甚至几百个；③非转录的间隔区短而且一致。

2）一类基因家族各成员的编码产物上具有大段的高度保守氨基酸序列。这对基因发挥功能是必不可少的。这些基因家族的各基因中有部分十分保守的序列。但家族成员间总的序列相似性却很低。

3）还有一类基因家族各成员的编码产物之间只有一些很短的保守氨基酸序列，从 DNA 水平上看，这些基因家族的成员之间的序列同源性更低。但其基因编码产物具有相同的功能，因为在蛋白质中存在发挥生物功能所必不可少的保守区域。

4）超基因家族（gene superfamily）中各基因序列间没有同源性，但其基因产物的功能相似。蛋白质产物中虽没有明显保守的氨基酸序列，但从整体上看却有相同的结构特征。如免疫球蛋白家族。

（二）基因外的重复 DNA 序列

除了基因家族外，染色体上还有大量无转录活性的重复 DNA 序列家族。与在基因家族中的组织形式类似，它们也有两种组织形式。一种是串联重复 DNA，成簇存在于染色体的特定区域。另一种是散布的重复 DNA，重复单位并不成簇存在，而是分散于染色体的各个位点上，它们也来源于 RNA 介导的转座（transposition）过程。散布的重复序列家族的许多成员是可转移的元件，不稳定的，可转移到基因组的不同位置。

1.卫星 DNA

(1)卫星 DNA　有些高度重复 DNA 序列的碱基组成和浮力密度同主体 DNA 有区别,在浮力密度梯度离心时,可形成不同于主 DNA 带的卫星带。卫星 DNA 的名称由此而来。卫星 DNA 由许多简单的重复单位组成。这些序列一般对应于染色体上的异染区域。也有一些高度重复序列的碱基组成与总体 DNA 相差不大,不能通过浮力密度梯度离心法分离(图 3-12)。

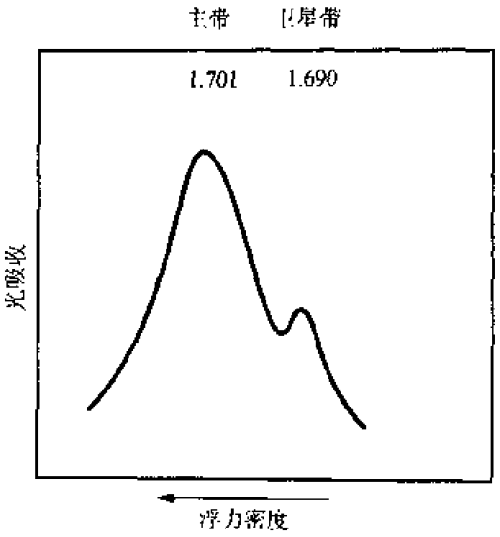


图 3-12　卫星 DNA

(2)卫星 DNA 的分类　根据重复单位的大小,这些非编码的高度重复的 DNA 序列可进一步分为卫星 DNA(satellite DNA)、小卫星 DNA(minisatellite DNA)、微卫星 DNA(microsatellite DNA)三类(表 3-4)。卫星 DNA 由长串联重复序列组成,一般对应于染色体上的异染区域。小卫星 DNA 由中等大小的串联重复序列组成,位于靠近染色体末端的区域,也可分散在核基因组的多个位置上,一般没有转录活性,其中有一些高变的小卫星 DNA。它们的重复单位之间的序列有很大不同。但都含有一个基本的核心序列:GGGCAGGAXG,多数靠近端粒,还有一些位于其他的位点。另一类小卫星 DNA 是端粒 DNA,它的主要成分是六核苷酸的串联重复单位 TTAGGG,作为一种缓冲成分,在真核生物染色体末端的复制中起重要作用。微卫星 DNA 是由更简单的重复单位组成的小序列,分散于基因组中,大多数重复单位是二核苷酸,也有少量含有三核苷酸和四核苷酸的重复单位。

表 3-4　人基因组的主要串联重复序列

分类	长度	重复单位大小(bp)	染色体定位
卫星 DNA	100 kb~数 Mb		
卫星序列 2 和 3		5	整个染色体
卫星序列 1		25~48	大多数染色体着丝粒
α		171	和其他异染色质区域 所有染色体着丝粒
β(Sau3A 家族)		68	1,9,13,14,15,21,22 号 和 Y 染色体的着丝粒
小卫星 DNA	0.1~20 kb		
端粒家族		6	所有染色体端粒
高变家族		9~24	所有染色体 通常靠近端粒
微卫星 DNA	小于 150 bp	1~4	所有染色体

(3)散布的重复DNA序列家族　在高度分散的重复DNA家族中含有少量转座元件，根据其大小不同，可分为短散布元件(short interspersed element，*SINE*)和长散布元件(long interspersed element，*LINE*)。比较典型的 *SINE* 是人的 *Alu* 重复序列家族(表3-5)。在其他哺乳动物中也有类似的序列，如鼠的 *B1* 家族。

Alu 家族成员众多，大约有几十万个，平均每 6 kb 就有一个，每个长度约 300 bp，在其序列中有 A C C T 序列，可被限制性内切核酸酶 *Alu* I 所切割，所以得名。*Alu* 家族的各个成员之间有很大的同源性，从 *Alu* 家族序列的长度和重复频率上看，*Alu* 序列都更像高度重复序列，但它们不同于高度重复序列的串联集中分布，而是广泛地分布在非重复序列之间。

表3-5　人类主要的散布的重复序列

种类	重复单位大小	拷贝数
Alu 家族	280 bp	700 000～1000 000
LINE-1 家族	全长 6.1 kb，平均长度 1.4 kb	60 000～100 000
MER 家族	几百 bp	共 100 000～200 000，每个家族 200～10 000
THE-1 家族	2.3 kb	10 000
HERV/RTLV 家族	全长 6～10 kb	几千个

第八节　细胞器基因组[14～16]

真核生物有两类细胞器能携带遗传物质，即线粒体(mitochondrion)，和叶绿体(chloroplast)。大多数动物细胞只有线粒体，植物细胞既有线粒体又有叶绿体。

线粒体和叶绿体 DNA 可编码自身所需的蛋白质及 tRNA 和 rRNA，但细胞器基因组并不编码全部所需的蛋白质，只编码其中的一部分，其余所需的蛋白质及酶均由核基因编码，在细胞质中的蛋白质合成系统合成后输入线粒体和叶绿体。细胞器膜不允许核酸分子通过，因此二者基因表达所需的 RNA 均由自身提供。而细胞器中的 mRNA 不能透过细胞器膜进入细胞质中，所以必须含有自己的蛋白质合成系统。

大多数细胞器 DNA 是环状的 DNA 序列。在某些低等真核生物中也有少量的线粒体 DNA 是线性分子。通常在一个细胞中有许多细胞器，每个细胞器中有其基因组的许多拷贝。所以每个细胞中有大量的细胞器 DNA。

叶绿体基因组比较大。在高等植物中通常为 140 kb，在低等真核生物中常常高达 200 kb。动物细胞线粒体基因组比较小，人、鼠、牛都是 16.5 kb。与核 DNA 相比，线粒体 DNA 的所占比例很少，只有不到 1%。酵母线粒体基因组很大，酿酒酵母为 84 kb，每个细胞 22 个线粒体，每个细胞器有 4 个拷贝。在生长的细胞中，线粒体 DNA 的比例高达 18%。植物线粒体 DNA 的大小差异很大，最小 100 kb。

线粒体有自己的蛋白质合成体系，其中 rRNA 和 tRNA 均由线粒体自身编码合成。

线粒体的 tRNA 比核基因编码的 tRNA 要小。其 RNA 聚合酶、氨酰- tRNA 合成酶和核糖体蛋白质均由核基因编码。尽管是由核基因编码的，但合成系统的蛋白质组分却是细胞器专用的，它们不同于细胞质中蛋白质合成系统的蛋白质成分。例如，叶绿体和线粒体中的蛋白质合成系统对于抑制细菌蛋白质合成的抗生素如红霉素、氯霉素很敏感；而细胞质的蛋白质合成系统对环己酰亚胺敏感。

表 3-6　人线粒体基因组与核基因组的协同作用

产物	线粒体基因组编码	核基因组编码
氧化磷酸化系统	13 亚基	>80 亚基
NADH 脱氢酶	7 亚基	>41 亚基
CoQ 还原酶		4 亚基
细胞色素 bc1	1 亚基	10 亚基
细胞色素 c 氧化酶	3 亚基	10 亚基
ATP 合成酶	2 亚基	14 亚基
蛋白质合成系统	2 rRNA 22 tRNA 13 mRNA	所有线粒体 核糖体蛋白(约 70)
其他线粒体蛋白质		所有线粒体蛋白质、如 DNA 聚合酶、RNA 聚合酶、其他酶、结构蛋白和转运蛋白

线粒体其他蛋白质的合成也常常由核基因组和线粒体基因组共同参与。在酵母线粒体中是 ATP 酶由两个单位 F_0 和 F_1 组成的复合体，其中跨膜因子 F_0 的三个亚基由线粒体基因组编码，而可溶性的 F_1 ATP 酶的 5 个亚基是由核基因组编码在细胞质里合成的。细胞色素 c 氧化酶的各亚基也是两个基因组共同编码。细胞色素 bc1 复合物的一个蛋白质亚基来源于线粒体，而 6 个亚基来源于细胞质。核糖体的小亚基也包含线粒体编码的蛋白。目前人线粒体的基因性质已经研究清楚(表 3-6)。

叶绿体基因组也编码与蛋白合成有关的 rRNA 和 tRNA，其 tRNA 与线粒体中的类似。叶绿体基因组可编码大约 50 种蛋白质，包括 RNA 聚合酶和一些核糖体蛋白。同样，叶绿体基因是由叶绿体中的系统转录和翻译的。

叶绿体的蛋白合成系统与细菌的很相似。许多核糖体蛋白质与大肠杆菌的同源。RNA 合成酶 α、β、β′亚基的基因与大肠杆菌的核心酶基因同源。但是编码 β′亚基的基因含有一个内含子。叶绿体的内含子分为两类，一类位于 tRNA 基因上，类似酵母核 tRNA 基因的内含子；一类位于编码蛋白的基因上，类似线粒体的内含子。

叶绿体的作用是进行光合作用，大多数叶绿体基因产物是类囊体膜的成分或与氧化还原反应有关。有些复合物与线粒体复合物一样，一部分亚基由细胞器基因组编码而另

一部分亚基由核基因组编码。在叶绿体中也有只由一个基因组编码的蛋白质。

在已鉴定了的叶绿体基因中，45 个基因的产物为 RNA，27 个基因的产物是与基因表达有关的蛋白，18 个基因编码类囊体膜的蛋白质，还有 10 个基因的产物与电子传递功能有关。

第九节　基因鉴定[17]

构建 DNA 物理图谱的基本目的是在特定染色体区域中对基因定位。当我们获得一段 DNA 序列的克隆时，如何确定它是基因序列或仅仅是没有编码功能的 DNA 呢？在理论上基因的 DNA 序列和非编码 DNA 序列有两个主要的区别：首先所有有活性的基因都能表达并产生 RNA 产物，主要是 mRNA，mRNA 可与 DNA 上的基因序列杂交。其次，基因的 DNA 序列通常是保守的，因为基因表达的产物是细胞发挥功能所必需的，在编码产物的序列上发生的突变往往对生物体不利，在自然选择的过程中会被除去，因此 DNA 的编码序列和调控序列与非编码序列相比在进化中十分保守。另外，在编码 DNA 中如果突变产生终止密码子将使其产物丧失生物功能而使突变基因在自然选择中被淘汰，而在非编码 DNA 中对应于终止密码子的 DNA 三联体不会被选择掉，因此基因序列与非编码 DNA 序列相比经常含有相对长的可读框。利用基因的这些特征，人们采用以下一些方法来确定在一段 DNA 序列上是否存在基因。

一、与 RNA 或 cDNA 杂交

因为基因表达将产生与基因本身序列互补的 RNA 产物，待测的 DNA 克隆作为杂交探针与从不同组织中提取的 mRNA 或总 RNA 进行 Northern 印迹（Northern blot）杂交，出现阳性结果则表明在克隆片段中可能存在基因，这时可用适当的 cDNA 文库作进一步筛选。但这个方法有两个缺点：首先，有些基因可能只在特殊的组织细胞中或特定的发育时期才大量表达，有时 RNA 产物或 cDNA 文库中可能无此基因的产物；另外，外显子序列在探针中的比例可能很低，导致杂交信号的强度降低，如果基因不是强烈表达，阳性杂交的信号可能不容易检测出来。

二、Zoo-blot 杂交

由于面临自然选择压力基因的编码序列在许多物种中都是保守的，相反非编码 DNA 序列相对快速地积累突变，在种间不是十分保守。Zoo-blot 利用保守基因的两个特性检测基因的存在，首先与不同种类的基因组样品进行 Southern 印迹杂交，产生阳性杂交信号的基因组 DNA 克隆可能含有在进化中十分保守的编码序列，然后检查能够杂交的序列中是否含有可读框。依靠这种特性，我们可分离纯化我们不太熟悉的但具有某些功能的基因。其具体过程首先从待测区域克隆出一个较短片段，以其为探针检测其他物种的相关的 DNA。如果在一些物种中发现相关片段，就对这些片段进行测序，详细考察是否有可读框和外显子与内含子连接的典型序列。如果有的话，就分离邻近的基因组区域。

如果这个区域看起来,像外显子,就鉴定整个基因,然后分离相应的 cDNA 和 mRNA,最终鉴定蛋白质产物。

三、外显子捕捉

另一个快速扫描基因组片段鉴定外显子的技术叫做外显子捕捉(exon trapping)。它所采用的载体含有强启动子和被内含子中断的两个外显子。当载体被转染到细胞中,便通过转录产生包含两个外显子序列的 RNA。载体的内含子序列中具有限制酶位点,用于插入待测区域的基因组片段。如果待测片段不包含外显子,在剪接图谱上不会发生变化,mRNA 含有与亲代载体相同的序列;如果基因组片段含有外显子,而且外显子两个侧翼有部分内含子序列,在 mRNA 加工时,这个外显子两侧的剪接位点将被识别,外显子的序列就被插入 RNA 中载体的两个外显子之间,表明捕捉到了外显子。

四、CG 岛鉴定

另外脊椎动物基因常常伴随着 CG 岛(CG island),可以利用这个性质鉴定克隆 DNA 中的基因。CG 岛是短的低甲基化的富含 GC 的序列,经常被发现于脊椎动物基因的 5′末端。CG 岛的(G+C)%超过 60%,并常常含有高浓度的 CG 二核苷酸,高出平均水平 10~20 倍。因此,CG 岛通常是一些稀有的内切酶的限制位点,它们在含有 1~2 个 CG 二核苷酸的富含 GC 的区域断裂。在基因组中这种限制位点的紧密成簇存在表明含有 CG 岛,然后进行 Southern 印迹可鉴定基因。但这种方法对没有连接 CG 岛的基因不适用。

五、DNA 序列的计算机分析鉴定

当对 DNA 克隆测序后,可利用计算机分析确定这些序列是否属于基因的一部分,有两种主要分析方法:

(1)同源分析　使所有三种可读框的核苷酸序列和相应的氨基酸序列与记录在数据库中的 DNA 和蛋白质序列相比较。如果出现序列间的明显符合,表明是与基因有关的序列。

(2)外显子预测　这个方法适用于大量 DNA 序列的检测。程序扫描 DNA 序列,检测是否含有外显子内含子连接处和剪接分支位点处的保守序列,以及是否存在相对长的可读框,鉴定可能的外显子。

六、扣除杂交

当有两个细胞型,其中一个表达某种 mRNA,另一个不表达。可利用扣除杂交(subtractive hybridization)技术分离该 mRNA。首先用靶细胞株的 mRNA 制备 cDNA 分子,然后 cDNA 与另一个细胞的 mRNA 进行完全的杂交,以除去靶细胞中非专一性的序列。

这个步骤从 cDNA 中除去了在两个细胞株中共同的所有序列，所以其专一性依赖于两个细胞之间的亲密程度。当除去与另一个细胞 mRNA 杂交的 cDNA 序列后，剩余 cDNA 的与靶细胞的 mRNA 杂交来确定它们代表编码序列。这些克隆含有专一性对应于靶细胞 mRNA 的序列，然后再用它们鉴定特定的基因。

上述基因鉴定方法有其局限性：例如有些基因的限制性表达图谱使其很难被鉴定；如果外显子的比例非常低，基因组克隆的杂交信号就会很弱。另外许多基因的表达是组织专一性的，而且不连接 CG 岛。

第十节　人类基因组计划[18~20]

各种基因定位的方法虽然使我们可以对基因组有一定的认识，但要对基因组有充分的理解，还必须实现基因定位的终极方式：即基因组全序列的测定。引人注目的人类基因组计划(human genome project，HGP)就是以测定人类基因组全序列为目标的巨大工作。HGP 与曼哈顿原子弹计划和阿波罗登月计划一起被称为本世纪三大科学工程。而 HGP 因其涉及面更广，影响力更大，又适逢世纪之交，成为承前启后、将贯穿于未来整个 21 世纪的最伟大的科学工程。人们在研究中发现，把一种生物学功能与一种或几种基因相对应起来的研究方法，并不能很好地阐明生物学功能内在的、真实的基因机制，其主要原因有两个：一是人类基因组所蕴含的约 10 万个基因迄今为止只有大约 10% 已被克隆或鉴定，对于与复杂生物学功能相关的基因可能只知道其中一部分，按照传统的研究模式，不能从整体上搞清生物学功能的基因机理；二是这些相关基因是通过相互作用实现生物学功能的，孤立的研究不可能全面了解基因间的相互作用。因此，人们认识到应进行基因组水平上的整体研究，提出只有先把人类基因组搞清楚，一切问题才有可能迎刃而解。围绕是否进行人类基因组计划，美国科学界展开了长达数年的争论，最终由国会批准于 1990 年 10 月正式启动 HGP，为期 15 年。

HGP 的主要目标是测定含有 30 亿 bp 的人类基因组全序列，为了实现这一目标，需进行以下工作：如构建遗传连锁图谱和物理图谱(physical map)，对模式生物基因组进行比较研究，建立新的技术方法，建立信息交流系统，对相关的社会、法律和伦理问题进行研究，技术培训等。1993 年，根据研究的实际进展，HGP 计划作了第一次调整，增加了基因鉴定的内容(即构建转录图谱)，这促进了基础理论研究和实际应用的结合。在随后的几年中，HGP 获得了惊人的进展，截至到 1998 年末，人们已得到了一些小生物体如大肠杆菌、酵母的全基因序列，对多细胞动物秀丽隐杆线虫的全序列测定也进入到了最后阶段。1998 年，HGP 在许多指标提前完成的情况下，再次调整战略目标，制定了下一个五年计划(1998~2003)，从计划中我们可以看到 HGP 提前实现的曙光(表 3-7)。在这个五年计划中，共有 8 个主要目标，以下是其详细内容：

一、人类 DNA 序列测定

HGP 的首要目标仍是提供公开的、完全的、高质量的人类基因组 DNA 序列。最近的

技术进步和大量测序的实践经验使人们有可能在2003年末之前完成高质量的人类基因组全序列测定，比最初计划提前两年完成。其详细内容见表3－7：

表3－7　HGP的计划完成情况和新的目标

领域	1993～1998计划	现状	1998～2003计划
遗传图谱	平均2～5 cM分辨率	在1994年发表1 cM图谱	已完成
物理图谱	建立30 000个STS的图谱	已建立52 000个STS图谱	已完成
DNA测序	完成80 Mb序列测定	完成180 Mb人类序列 和111 Mb非人类序列测定	在2001年底完成1/3人类序列 测定和剩余的工作草图 在2003年完成人类全序列测定
测序技术	改进和创新	每年90 Mb 每个碱基0.5美元 使用毛细管阵列电泳技术	每年500 Mb， 每个碱基少于0.25美元
人类序列变异研究			100 000 SNP图谱， 开发新技术
基因鉴定	开发新技术	30 000 EST图谱	全长的cDNA
功能分析			开发基因组水平的研究技术
模式生物	大肠杆菌全序列 酵母全序列 秀丽隐杆线虫大部分序列 开始果蝇测序 建立10 000 STS小鼠图谱	在1997年发表 在1996年发表 80%完成 完成9% 建立了12 000个STS图谱	在1998年底完成 在2002年完成 开发其他基因组来源 在2005年完成序列测定 在2005年前建立工作草图

1）在2001年底前，完成1/3人类DNA序列测定；
2）在2001年底前，在基因组文库基础上绘制出基因组的草图，包括至少90%基因组内容，但草图中仍有许多DNA区域是中断的，而且相互间的顺序仍不明确；
3）使序列进一步完整，并对公众公开；
4）在2003年底前，完成人类基因组全序列测定。

二、测序技术的发展

1）继续提高测序反应的效率并降低当前测序技术的费用；
2）支持新技术的研究，使测序技术得到明显的改进；
3）寻找有效的方法使新技术得到大规模的实际应用。

三、人类基因组序列变异的研究

序列变异是所有基因组的基本性质，人类基因组表现出大量的位点和形态的不同(多

态性，polymorphism)。人类基因组最常见的多态性是单碱基差异，又叫做单核苷酸多态性(single nucleotide polymorphism，SNP)。SNP 在基因组中大量存在、广泛分布，它是鉴定肿瘤、糖尿病、精神病等人类疾病的重要指标。因此 SNP 的研究是鉴定致病基因的有效方法。

1）发展快速、大规模的鉴定和寻找 SNP 与其他 DNA 序列变异的技术；

2）在今后的 5 年中，在大多数已发现基因的编码区内鉴定基本的变异；

3）建立至少含有 100 000 个标记的 SNP 图谱；

4）序列变异的基础理论研究；

5）建立 DNA 样品和细胞系的公共资源。

四、功能遗传学的研究方法

DNA 的序列和结构研究仅仅是对基因组的初步了解，进一步的工作是阐明 DNA 序列的功能。当前在基因组水平研究 DNA 功能的方法包括通过序列比较分析来推测功能，分析基因的 mRNA 和蛋白质产物，以及一些基因断裂的方法。人类基因组计划通过对人类所有基因的全长 cDNA 作鉴定并测序，并进一步鉴定其转录产物和蛋白质产物的性质，进行基因功能的分析。

1）扩展 cDNA 资源；

2）发展非蛋白质编码序列 DNA 功能的研究方法；

3）开展基因表达的全面分析技术的研究；

4）改进基因组水平的诱变技术；

5）球蛋白分析技术的研究。

五、比较遗传学研究

因为所有生物体在基本进化途径上是相关的，所以对其他模式生物体的比较研究有助于人类基因组的研究。人们在基因组计划的第一个五年内已测定了大肠杆菌和酵母的全序列。今后的工作包括：

1）在 1998 年完成秀丽隐杆线虫的全序列测定；

2）在 2002 年完成果蝇基因组的全序列测定；

3）鼠基因组的研究计划包括，继续进行其基因组物理图谱和遗传图谱的研究，扩展 cDNA 资源，在 2005 年完成全序列测定；

4）开展对其他模式生物的研究，以加深对人类基因组的了解。

六、伦理、法律和社会问题的研究

详细内容见其他有关专著。

七、生物信息学和计算机生物学

人类基因组计划需要利用计算机技术的帮助，建立 DNA 序列数据库并提供分析工具。

1)提高数据库的容量和用途；

2)发展数据产生、捕获、注释的计算工具；

3)建立并发展用于比较功能研究的工具和数据库；

4)发展并改进用于描述和分析序列相似性和差异性的工具；

5)建立软件共享的机制。

八、遗传学研究专门人才的教育训练

详细内容见其他有关专著。

（吴东林）

参 考 文 献

[1] Lewin B. Genes, NewYork:Oxford University Press. 1997

[2] Ingram V M. Nature, 1957. 180:326～328

[3] Crick F H et al. Nature, 1961. 192:1227～1232

[4] Breathnach R et al. Ann Rev Biochem, 1981. 50:349～383

[5] Sanger F et al. Nature, 1977. 265:687～695

[6] Goebl M G et al. Cell, 1986. 46:983～992

[7] Geraghty D E et al. J Immunol, 1992. 149:1934～1936

[8] Weiner A M et al. Ann Rev Biochem, 1986. 55:631～661

[9] Swallow D M et al. Nature, 1987. 328:82～84

[10] Korenberg J R et al. Cell, 1988. 53:391～400

[11] Southern E M. J Mol Biol, 1975. 94:51～70

[12] Jeffreys A J et al. Nature, 1985 314:67～73

[13] Neer E J et al. Nature, 1994. 371:297～300

[14] Anderson S et al. Nature, 1981. 290:457～465

[15] Palmer J D. Ann Rev Genet, 1985. 19:324～354

[16] Poyton R O et al. Ann Rev Biochem, 1996 65:563～608

[17] Antequera F. et al. Proc Natl Acad Sci USA, 1993. 90:11995～11999

[18] Collins F et al. Science, 1983. 262:43～46

[19] Deloukas P et al. Science, 1998. 282:744～746

[20] Collins F S et al. Science, 1998. 282:682～689

第四章　复　　制

DNA 是遗传信息的载体，生物体要保持物种的延续，子代必须从亲代继承控制个体发育的遗传信息。生物体通过细胞分裂来实现物种的传递和个体数目的增殖，在此过程中细胞染色体的数目加倍，作为遗传物质的 DNA 分子的数目也加倍（复制），并伴随染色体平均分配到两个子代细胞中。细胞分裂有两种形式，有丝分裂（mitosis）与减数分裂，二者的差别在于，前者 DNA 复制一次，细胞分裂一次，后者 DNA 复制一次，细胞连续分裂两次，子代染色体数目减半。在细胞分裂过程中，如果 DNA 的含量保持不变，每个子代细胞中遗传物质的含量将随着连续不断的分裂循环而稀释甚至消失。所以，DNA 分子的复制在保持生物物种遗传的稳定性方面起着重要的作用。复制（replication）是指以原来的 DNA 分子为模板合成出相同分子的过程。遗传信息通过亲代 DNA 分子的复制传递给子代。

无论是只含有一条染色体的原核生物还是含有多条染色体的真核生物，每次发生有丝分裂，细胞内的基因组都精确地复制一次。复制与细胞周期是什么关系呢？DNA 复制是细胞分裂的决定因素，直到 DNA 复制完成前，细胞都不会发生分裂，而 DNA 复制的完成恰恰是细胞分裂的触发点，复制后的基因组伴随细胞分裂被平均分配到每个子细胞中。细胞周期调控基因控制 DNA 的复制并触发细胞分裂。在原核生物中，DNA 复制是由细菌染色体上一个特殊位点开始的，新细胞膜的生长意味着细胞分裂过程的完成。在真核细胞中，复制起始于前 S 期，DNA 的合成发生于 S 期，一系列的调控事件控制细胞进行随后的有丝分裂过程。

第一节　复 制 概 况

一、半保留复制

在复制过程中，DNA 的数量加倍。在复制结束后，细胞分裂前，细胞中的 DNA 分子由复制前的亲代 DNA 和复制后的产物组成，DNA 分子有两个拷贝。

DNA 分子的碱基互补配对原则是 DNA 分子结构的基础，这个原则在 DNA 复制过程中也起着重要的指导作用。可推测，DNA 分子复制是这样进行的：首先，碱基间的氢键断裂使 DNA 分子双链解旋，分开；然后，每条单链作为模板合成其互补链，各自形成一个新的双链 DNA 分子。新合成的两个子代 DNA 分子与原来的亲代 DNA 分子碱基序列完全一样。每个子代分子都有一条单链来自亲代 DNA 分子，另一条单链是新合成的。两个子代 DNA 分子通过细胞分裂平均分配到两个子代细胞中。由于每个子代细胞中只有一半的遗传物质是来自亲代细胞，这种复制方式被称为半保留复制（semiconservative replication）。

理论上 DNA 复制也可能采取其他方式，如全保留方式，甚至混合复制方式（图 4-1）。

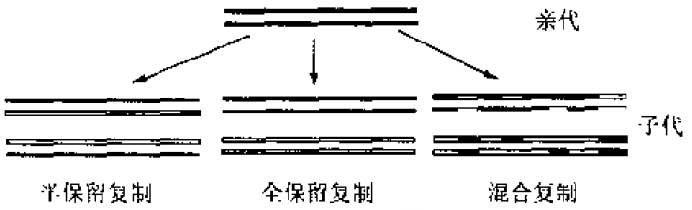


图 4-1 子代 DNA 分子的可能复制方式

真实情况究竟如何呢？1958 年，Meselson 和 Stahl 设计了一个很巧妙的实验证明了 DNA 复制是采取半保留机制。他们首先让大肠杆菌在以$^{15}NH_4Cl$ 为唯一氮源的培养基中生长，经过多代培养所有的 DNA 分子都被标记上^{15}N。^{15}N-DNA 的密度比普通^{14}N-DNA 的大，通过氯化铯密度梯度离心，两种 DNA 形成位置不同的区带。将有^{15}N 标记的大肠杆菌转移到普通的培养基（含^{14}N 的氮源）中培养，一代后，所有 DNA 的密度都介于^{15}N-DNA 和^{14}N-DNA 之间，即形成^{15}N-^{14}N 杂合分子。两代后，^{14}N 分子和^{15}N-^{14}N 杂合分子等量出现。继续培养，^{14}N-DNA 增多，说明在 DNA 分子复制时，原来的 DNA 分子被分为两个亚单位，分别构成子代分子的一半。证明了基因的半保留复制机制（图 4-2）。

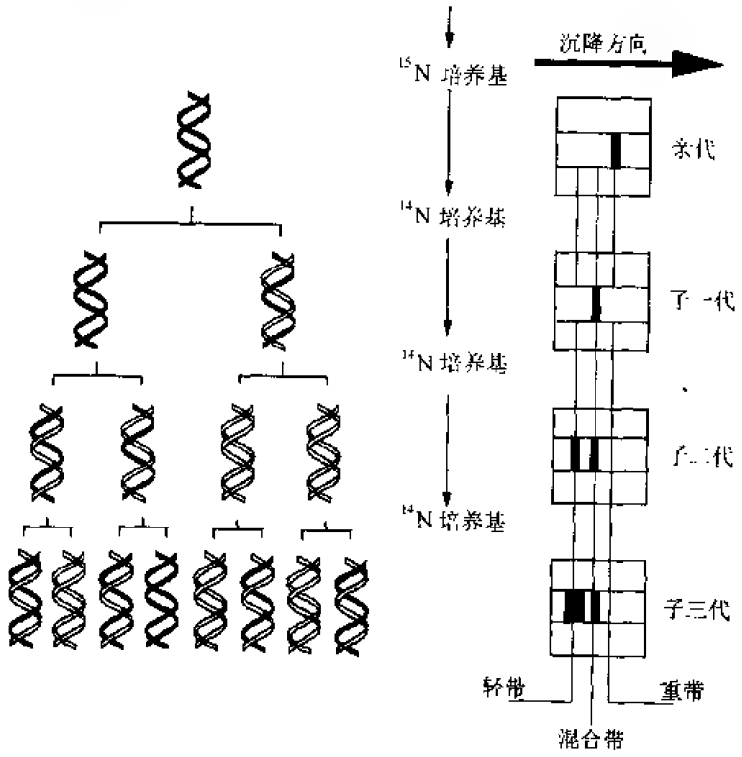


图 4-2 半保留复制的证明

半保留复制是双链 DNA 普遍采用的复制机制。即使单链 DNA 分子，在其复制过程中通常也要先形成双链的复制形式。半保留复制要求亲代 DNA 分子的两条链分开，各自作为模板，通过碱基配对原则，合成另一条互补链。模板就是能提供合成一条互补链所需信息的核酸链。半保留复制机制说明了 DNA 在代谢上的稳定性。但实际过程并不像

上述这么简单。半保留复制作为一个复杂的过程，需要多种细胞组分的参与；受到细胞中多种条件的控制；同时为保持遗传的稳定性，要具有高度的忠实性，否则随着不断复制，将产生严重的差错，从而影响生物体个体的存活。

二、复制起点、方向和方式

由很多体系用不同的实验手段获得的结果都证明，原核生物的复制是在 DNA 分子的特定位点开始的，这一位点被叫做复制起点，常用 *ori* 表示。原核生物的染色体只有一个复制起点，复制从起点开始，进行到终点结束，完成整个染色体 DNA 分子的复制，即原核生物的染色体只有一个复制单位。真核生物的实验表明，复制也从特定的位点开始，以双向延伸的方式进行，直到终点为止，但其复制是从多个位点开始的，所以真核生物的 DNA 分子上有多个复制单位。每一个这样的 DNA 单位称为复制子(replicon)。复制子是基因组中能独立进行复制的单位。

表 4-1　不同生物的复制子

种类	复制子数目	复制子长度(kb)	复制叉移动速度(bp/分)
细菌	1	4200	50 000
酵母	500	40	3600
果蝇	3500	40	2600
蟾蜍	15 000	200	500
小鼠	25 000	150	2200
植物	35 000	300	

在一个细胞分裂周期中，每个复制子只启动一次。每个复制子都有控制复制开始的复制起点(origin)，可能还有终止复制的终点(terminus)。任何起点与终点之间的序列，都作为复制子的一部分参与复制。起点是一个反式作用位点，只影响其所在的那部分 DNA。为实现适当的遗传，细菌的复制子必须具有下列功能：起始一个复制循环；控制起始事件的频率；把完成复制的染色体分配到子细胞中去。前两个功能是由起点处的序列控制的；分配是一个独立于复制的功能，但事实上仍与起点附近的序列有关。通过对功能缺失的分配突变型的鉴定，可确定与分配功能有关的序列。真核生物的起点没有分配功能，只与复制有关。

原核生物细胞的基因组只有一个复制子，所以复制和分裂的单位是一致的，最大的复制子就是细菌的染色体本身。整个基因组的复制起始于单一位点(起点)，每次细胞分裂基因组复制一次。每个单倍体细菌只有一个染色体，这样的复制调控叫单拷贝。除了染色体，细菌还可能含有质粒。质粒是一个自主复制的环状 DNA 基因组，构成一个独立的复制子。质粒复制子可能与细菌基因组一同复制，显示单拷贝控制；质粒的拷贝数也可能比细菌基因组的拷贝数多，这时质粒的复制就受多拷贝控制的复制。噬菌体和病毒的 DNA 也是单复制子，但它们在一个感染周期中可以起始许多次。与单拷贝控制型的染色体相反，叶绿体和线粒体中染色体的调控更类似于细菌质粒的多拷贝控制。每种细胞器

DNA 在一个细胞内有多个拷贝，而且细胞器 DNA 的复制调控与细胞周期相关。

如图 4－3 所示，参与复制的 DNA 分子上有两个区域，未复制的区域由亲代 DNA 组成，已复制的区域由两条子代链组成。复制正在发生的位点叫做复制叉（replication fork）或生长点（growing point），复制叉从起点开始沿着 DNA 移动。

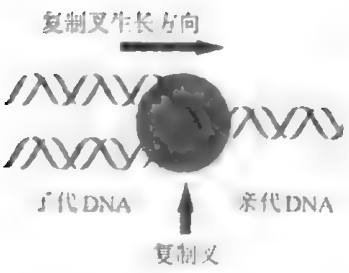


图 4－3　复制叉的结构

复制可单向进行，也可双向进行，这取决于在起点有一个还是两个复制叉。单向复制时，有一个复制叉从起点开始沿着 DNA 移动；双向复制时，两个复制叉沿着相反的反向远离起点（图 4－4）。在电子显微镜下观察正在复制的 DNA，复制的区域形如一只眼睛，但无法分辨是单向复制还是双向复制。如果是单向复制，这个复制眼（replicative eye）由一个固定的起点和一个移动的复制叉组成；如果是双向复制，这个复制眼代表一对复制叉。两种情况下，复制都随着眼的扩张最终扩展到整个复制子。环状 DNA 的复制眼可形成如图 4－5 那样的 θ－结构。

大多数生物体内 DNA 的复制都以双向等速方式进行，但也有例外情况。例如在枯草杆菌中，复制是从起点开始双向进行的，但两个复制叉的移动是不对称的，在一个方向上移动染色体 1/5 的距离，然后停下来等另一个方向复制 4/5 的距离。质粒 ColEI 的复制完全是单向进行的。大多数生物体的 DNA 复制都以对称方式进行，即 DNA 分子中的两条链同时复制；但也有在一定时期内 DNA 只复制一条链的情况，比如后面要讲到的线粒体的 D 环复制和一些细菌噬菌体的滚环复制方式。

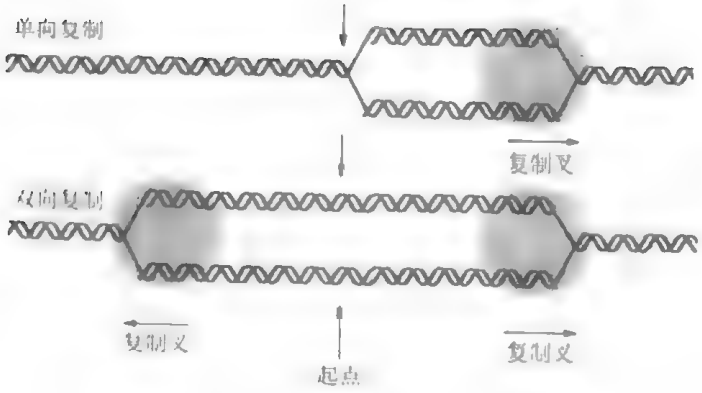


图 4－4　DNA 的单向或双向复制

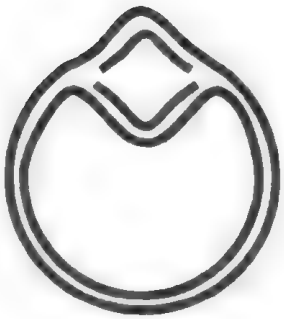


图 4－5　环状 DNA 的复制眼形成 θ－结构

用遗传的方法可以证明大肠杆菌的 DNA 复制是从一个特定位点开始双向等速进行。细菌染色体复制一次的时间相当于细胞繁殖一代所需的时间。如果每个细菌从起点同时开始复制，在一定的时刻，靠近复制起点的基因复制得多，而远离复制起点的基因就复制得少些。如果进行单向复制，基因图上基因频率将呈单向梯度，最高频率的基因和最低频率的基因是连锁的；如果是双向复制，基因频率将以起点为中心呈双向梯度。测定发现，基因频率是以 *ilv* 基因为中心双向下降，这说明大肠杆菌以 *ilv* 基因附近为复制起点开始双向复制。

当一段来自起点的 DNA 序列与缺失起点的 DNA 分子克隆到一起时，如果从起点来

的DNA片段包括一个复制起点所必须具备的序列，它将重建一个具有自主复制功能的质粒，支持与其连接的任何DNA序列的复制。由此可鉴定起点的DNA序列。细菌、酵母以及叶绿体、线粒体的DNA复制起点现均已被克隆，其核苷酸序列也被确定了，它们共同的特点是富含A-T序列。这可能与复制起始时在起点处DNA双螺旋分子的解旋有关。大肠杆菌的基因组是从*oriC*位点开始双向复制，含有*oriC*的人工质粒可在大肠杆菌中复制并存活下来，通过不断缩小人工质粒中克隆的*oriC*片段的大小，最终确定开始复制需要245 bp的区域。

细菌的染色体是环状的，两个复制叉从起点开始以相反方向沿着染色体移动，在起点对面的中间点相遇并停止复制。但复制不是通过复制叉相遇而简单地终止的。在这个相遇点两边100 kb的区域中各有两个终止区域（*terE*、*D*、*A*和*terF*、*C*、*B*），每个终点只对特定方向的复制叉有终止作用。所有*ter*序列都包含一段短的（22 bp）区域，可在体外终止复制。终止的实现需要*tus*基因的产物，这种蛋白质识别*ter*共有序列并阻止复制叉行动。

所以，细菌染色体作为一个单位从一个特殊的起点开始双向复制，复制终止在独立的区域，随后在识别DNA高级结构的酶的帮助下两个子染色体分开。

三、DNA聚合酶与DNA聚合反应

复制是由DNA聚合酶（DNA polymerase）负责完成的，从DNA聚合酶的字面上可以很清楚地知道它具有在模板链上催化新DNA链合成的活性。在适量DNA和镁离子存在下，DNA聚合酶催化4种脱氧核糖核苷酸合成DNA，每次在DNA的3′ OH末端加入一个核苷酸使DNA链由5′向3′方向延长，同时释放无机焦磷酸。合成的DNA与天然的DNA有相同的化学结构和物理化学性质（图4-6）。DNA聚合反应可表示如下：

$$\text{poly(核苷酸)}_n\text{-OH3}' + \text{dNTP} \rightarrow \text{poly(核苷酸)}_{n+1}\text{-OH3}' + \text{PPi}$$

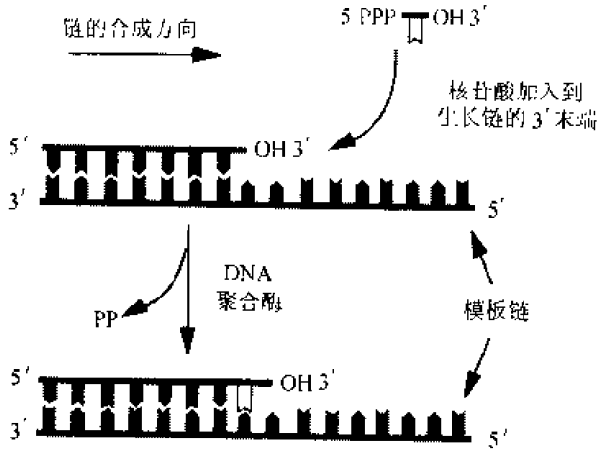


图4-6 DNA聚合酶催化的链延长反应

DNA聚合酶催化反应是按模板的指令进行的。只有加入的碱基与模板链上的碱基

形成 Watson-Crick 类型的碱基配对时，才能在聚合酶的催化下形成磷酸二酯键。因此，DNA 聚合酶是一种模板指导的聚合酶，产物 DNA 与模板的碱基组成相同，这说明在 DNA 聚合酶的作用下，模板 DNA 的两条链都能复制。

四、复制的基本过程

DNA 的复制包括起始、延伸、终止三个步骤。复制子增殖依赖于起始过程中起点处发生的相互作用，复制调控发生在起始阶段。一般复制开始后，就会进行到底，直到整个基因组都复制完毕。起始频率由复制起点和调控蛋白二者控制。DNA 聚合酶把新生链的第一个核苷酸加到引物的 3′羟基上，从而开始新生链的延伸合成。在单方向复制的环状分子中，复制终点就是它的复制起点。而在双方向复制的环状分子中，有的有固定的终点，而大多数没有固定的终点，只是两个生长点的简单碰撞。环状分子复制的最后一步，是环连分子的分离，这是由拓扑异构酶催化的反应。

五、DNA 复制的半不连续性

DNA 分子的两条链是反向平行的，这种结构产生了复制上的问题。当复制叉前进时，在两个暴露的亲链上都要进行子链的合成。在一条链上复制叉以 5′→3′方向移动，在另一条链上则以 3′→5′方向移动。但是所有已知的 DNA 聚合酶的合成方向都是 5′→3′，即新生链的生长方向只能是 5′→3′，而不能相反。这就很难解释 DNA 在复制时两条链如何能够同时作为模板合成其互补链。为了解决这个矛盾，冈崎等人提出了 DNA 的不连续复制模型，认为 3′→5′走向的 DNA 实际上是由许多 5′→3′方向合成的 DNA 片段连接起来的。

1968 年，冈崎等用^{3}H－脱氧胸苷标记噬菌体 T4 感染的大肠杆菌，然后通过碱性密度梯度离心法分离标记的 DNA 产物，发现在短时间内首先合成的是较短的 DNA 片段，接着出现较大的放射性标记的分子。最初出现的 DNA 片段长度为 1000～2000 个核苷酸(nt)，称为冈崎片段(Okazaki fragment)。用 DNA 连接酶的温度敏感突变体进行实验，在连接酶不起作用的温度下，便有大量的 DNA 片段的积累。这些实验说明在 DNA 复制过程中首先合成较短的片段，然后再由连接酶连接成大分子 DNA。

冈崎等最初的实验不能断定 DNA 链的不连续合成只发生在一条链上，还是在两条链上都如此。对冈崎片段进行测定，结果测得的数量远超过新合成 DNA 的一半，似乎两条链都是不连续复制的。后来发现这是由于尿嘧啶取代胸腺嘧啶掺入 DNA 所造成的。错误掺入的尿嘧啶可被尿嘧啶 *N* - 糖苷酶切除，产生无尿嘧啶的磷酸二酯键，这种磷酸二酯键很容易被碱水解而断裂，从而形成一些较短的片段。这种情况下的冈崎片段应该称为尿嘧啶片段，它并不能反映不连续合成的真实情况。当用缺乏糖苷酶的大肠杆菌突变体(*ung*)进行实验时，DNA 的尿嘧啶将不再被切除。此时，新合成的 DNA 大约有一半放射性标记出现子冈崎片段中，另一半直接进入大的片段。由此可见，当 DNA 复制时，一条链是连续的，另一条链是不连续的，因此称为半不连续复制(semidiscontinuous replication)(图 4－7)。

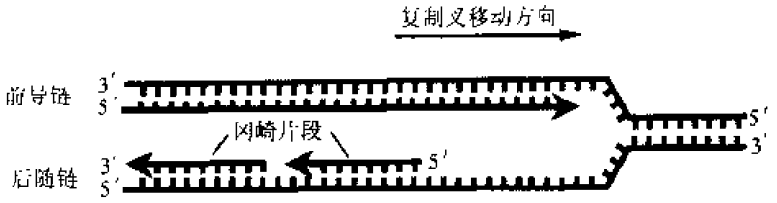


图 4-7 DNA 的半不连续复制

以复制叉向前移动的方向为标准，一条模板链是 3′→5′走向，在其上 DNA 以 5′→3′方向连续合成，称为前导链(leading strand)；另一条模板链是 5′→3′走向，在其上 DNA 也是以 5′→3′方向合成，但是与复制叉移动的方向相反，所以随着复制叉的移动，形成许多不连续的片段，最后连接成一条完整的 DNA 链，该链称为后随链(lagging strand)。

表 4-2 参与大肠杆菌复制的基因

基因	基因产物及功能
dnaA	大肠杆菌复制起始蛋白，分子量 52 kDa，聚集于 *oriC*
dnaB	六聚体螺旋酶，单体分子量 52 kDa，参与引发反应和链延伸
dnaC	引发体组分，分子量 25 kDa，六个单体结合 DnaB 六聚体，是 DnaB 的装载和改型因子
dnaE	PolIII 全酶的 α 亚基，分子量 130 kDa，具有聚合和校对功能
dnaF	核糖核苷酸还原酶，又称为 RndA
dnaG	引发酶，合成 RNA 引物，分子量 64 kDa
dnaN	PolIII 全酶的 β 亚基，单体分子量 40 kDa。以二聚体形式存在，是 PolIII 的进行性因子
dnaQ	曾叫做 MutD，编码 PolIII 的 ε 亚基，分子量 27 kDa，是起校正作用的 3′→5′外切核酸酶
dnaT	编码产物 22 kDa，又叫做蛋白 i 和因子 x
dnaX	编码 PolIII 的 τ 亚基和 γ 亚基，分子量分别为 71 和 48 kDa。τ 亚基有 ATPase 活性，可使核心酶二聚化，γ 亚基参与形成 γ 复合物
dut	dUTPase
gyrA	DNA 旋转酶的 α 亚基
gyrB	DNA 旋转酶的 β 亚基
holA	PolIII 全酶的 δ 亚基，单体分子量 39 kDa，参与形成 γ 复合物
holB	PolIII 全酶的 δ′亚基，单体分子量 37 kDa，参与形成 γ 复合物
holC	PolIII 全酶的 χ 亚基，单体分子量 17 kDa，参与形成 γ 复合物
holD	PolIII 全酶的 ψ 亚基，单体分子量 15 kDa，参与形成 γ 复合物
holE	PolIII 全酶的 θ 亚基，单体分子量 8.6 kDa，参与形成 γ 复合物
hu	刺激依赖 *oriC* 的 DNA 复制

续表

基因	基因产物及功能
lig	DNA 连接酶，分子量 75 kDa
polA	大肠杆菌 PolI，冈崎片段的连接
priA	引发体组分，又称为 n′，分子量 81 kDa
priB	引发体组分，又称为 n，分子量 12 kDa
priC	引发体组分，又称为 n″分子量 21 kDa
rep	螺旋酶
rpoA、*B*、*C*、*D*	RNA 聚合酶亚基，复制起始，在某些体系中合成引物
ssb	单链 DNA 结合蛋白，在复制中保持单链稳定性
tus	复制终止蛋白，结合 *ter* 序列，分子量 36 kDa
ung	尿嘧啶- N -糖基酶，从 DNA 上除去尿嘧啶

第二节　复制体系

一、复制体系的鉴定

通常人们通过对突变体表型的变化来考察基因的功能，如果某个基因突变导致细胞不能复制，就可以说这个基因与复制有关。由于缺乏复制 DNA 的能力对细胞来说是致命的，所以我们能够得到的 DNA 复制突变体只能是条件致死突变体，它在合适的条件下可以完成复制，在非允许条件下（如高温，42℃）则不能进行 DNA 复制。利用大肠杆菌的温度敏感突变体已鉴定了一系列的 *dna* 基因座。根据 *dna* 基因突变体在温度升高时表现出的性质，可将它们分为两类：一类是快停突变体（fast-stop mutant），当温度升高时，复制立即停止，它们缺失复制体系组分，特别是延伸所需的酶或参与供应关键前体的酶；另一类是慢停突变体（slow-stop mutant），在非允许温度下，还可完成当前的复制，但不能起始新的复制，它们的产物是从起点起始复制所必需的。用温度敏感突变体鉴定出来的 *dna* 基因产物并非都直接参与 DNA 复制，但都与复制过程有关。如快停突变型 *rndA* 原来叫做 *danF*，该基因编码核糖核苷酸还原酶的一个亚基，与供应复制所需的前体有关。

历史上，对参与 DNA 复制的蛋白质进行的识别和鉴定沿用二种方法：纯化、重建和突变，三种方法往往一起使用。体外互补实验是鉴定复制体系组分的重要手段，方法是从 DNA 突变体中分离出复制体系，此时突变基因的产物是无活性的，然后在非允许条件下加入相应野生型的抽提物，观察复制活性的恢复，通过鉴定抽提物中的活性组分确定 *dna* 基因编码的蛋白质。表 4 - 2 列出了参与大肠杆菌复制的部分基因。

二、DNA 聚合酶

上一节我们已提到了 DNA 聚合酶催化的聚合反应。原核生物和真核生物中都含有

多种 DNA 聚合酶，只有一部分真正地参与了复制反应，这些酶被称为 DNA 复制酶。其他酶参与了复制的附属反应或损伤 DNA 的修复合成。

1. 大肠杆菌的 DNA 聚合酶

在大肠杆菌中发现了三种 DNA 聚合酶：DNA 聚合酶 I、DNA 聚合酶 II 和 DNA 聚合酶 III。DNA 聚合酶 I 参与损伤 DNA 的修复，并在半保留复制中起辅助作用。DNA 聚合酶 II 也参与修复。DNA 聚合酶 III 是一个多亚基的蛋白质，是合成 DNA 新链的主要复制酶。表 4－3 概括了三种酶的性质。

(1) DNA 聚合酶 I　大肠杆菌 DNA 聚合酶 I 是第一个被鉴定的 DNA 聚合酶，它由 *polA* 基因座编码，是一个 103 kDa 的单链多肽。每个大肠杆菌细胞中约有 400 个分子的 DNA 聚合酶 I。当底物和模板存在时，DNA 聚合酶 I 可使脱氧核糖核苷酸依次加到具有 3′－OH 末端的多核苷酸链上。它的催化需要引物链(DNA 或 RNA 链)的存在。在 37℃ 条件下，每分子 DNA 聚合酶 I 每分钟可催化约 1000 个核苷酸的聚合。DNA 聚合酶 I 是一个多功能酶，它可以催化以下的反应：①DNA 链沿 5′→3′方向延长(DNA 聚合酶活性)；②从 3′端水解 DNA 链[3′→5′外切核酸酶(exonuclease)活性]；③从 5′端水解 DNA 链(5′→3′外切核酸酶活性)。用蛋白水解酶将 DNA 聚合酶 I 作有限水解，可得到分子量为 68 kDa 和 36 kDa 的两个片段。较大的断裂产物(68 kDa)叫 Klenow 片段，在体外它可催化合成反应，Klenow 片段有 DNA 聚合酶活性和 3′→5′外切核酸酶活性，两种活性分别处于片段的不同区域。C 端的 2/3 片段具有聚合酶活性，N 端的 1/3 片段有外切核酸酶活性，在蛋白质中，两个活性位点的距离为 3 nm，表明碱基的加入和去除功能在空间上是分开的；小片段(35 kDa)有 5′→3′外切核酸酶活性，可以切除少量的核苷酸，一次最多切 10 个碱基。该酶在切除由紫外线照射而形成的嘧啶二聚体中起重要作用，在 DNA 的半不连续合成中，冈崎片段 5′端 RNA 引物的切除可能也依赖于该外切酶，小亚基的外切酶活性和大亚基的聚合校对功能协同作用，使 DNA 聚合酶 I 在体外具有从切口处起始复制的独特能力，而其他 DNA 聚合酶不具备这种能力。在双链 DNA 的磷酸二酯键断裂处，DNA 聚合酶 I 可延伸 3′末端，合成新的 DNA 片段后，DNA 聚合酶 I 可置换双螺旋中已有的同源链。该过程如图 4－8 所示。置换出的同源链被 5′→3′外切核酸酶降解。除部分片段被新合成的材料取代且缺口位置沿双螺旋移动外，DNA 的性质没有任何改变。该反应在实践中很有意义，体外的切口移位(nick translation)是实现在 DNA 分子上引入放射性标记核苷酸的重要技术。在体内，5′→3′聚合酶合成活性和 3′→5′外切核酸酶活性主

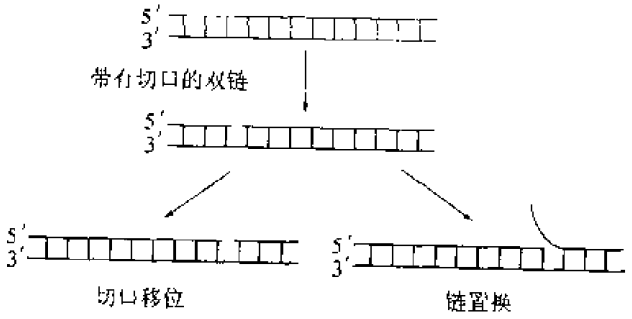


图 4－8　切口移位和链置换

要用来填充DNA双链中短的单链区域，复制过程中或从DNA中切除受损伤的碱基后会产生这些单链区域。

表4-3 大肠杆菌的DNA聚合酶

	DNA聚合酶 I	DNA聚合酶 II	DNA聚合酶 III
分子结构	单亚基	单亚基	全酶由10种共18个亚基组成
分子量(kDa)	103	120?	850
分子数/细胞	400	100	10～20
5′→3′聚合酶活性	+	+	+
37℃时的转化率(核苷酸数/酶分子.分)	1000	50	750～1000
3′→5′外切核酸酶活性	+	+	+
5′→3′外切核酸酶活性	+	−	−

(2) DNA聚合酶II　DNA聚合酶II是一条分子量120 000的多肽链。这个酶的活性很低，只有DNA聚合酶I的5%。它也以4种脱氧核糖核苷酸为底物，从5′→3′方向合成DNA。DNA聚合酶II具有3′→5′的外切核酸酶活性，但无5′→3′外切酶活性。它在DNA的修复中起一定作用。

(3) DNA聚合酶III　尽管在细菌中存在大量的DNA聚合酶I，但它并不是复制的主要酶。在*polA*的突变型中，复制可继续进行。目前，DNA聚合酶III(polymerase III，PolIII)被认为是真正负责大肠杆菌细胞内合成DNA的复制酶(replicase)。DNA聚合酶III是多亚基组成的蛋白质，它在细胞中含量很少，在每个细胞中只有10～20个拷贝的全酶(holoenzyme)。PolIII全酶由α、β、γ、δ、ε、θ、τ、δ、χ、ψ 10种不同的亚基组成，至少含3种重要的酶活性。全酶的亚基列于表4-4，它们可能的装配模型如图4-9所示。

表4-4 DNA聚合酶III全酶的亚基组成

亚基	基因	分子量(kDa)	功能及亚组装		
α	*dnaE*	129.9	DNA聚合酶	核心酶	PolIII*
ε	*dnaQ*，*mutD*	27.5	3′-5′外切核酸酶		
θ	*holE*	8.6	促进ε亚基的作用PolIII′		
τ	*dnaX*	71.1	使核心酶二聚化 结合γ复合物		
γ	*dnaX*	47.5	结合ATP(依赖DNA的ATP酶)	γ复合物	PolIII*
δ	*holA*	38.7	结合β亚基		
δ′	*holB*	36.9	结合γ和δ亚基		
χ	*holC*	16.6	结合SSB		
ψ	*holD*	15.2	结合χ和γ亚基		
β	*dnaN*	40.6	滑动钳**		

* PolIII*是PolIII的一种前体；　** sliding clamp.

从表4-4中可以看出，全酶含有4个不同的功能组分，以下分述之。

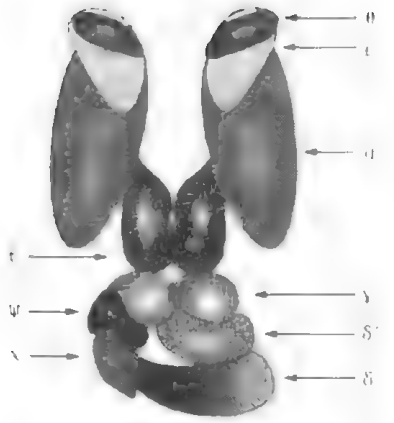


图 4-9 全酶结构模型

(1) 核心聚合酶(core polymerase) 核心聚合酶由 α、ε、θ 三种亚基组成,其中 α 亚基的分子量为 130 kDa,由 *dnaE* 编码合成,具有 5′→3′方向合成 DNA 的催化活性,每秒可以合成 8 个核苷酸。ε 亚基具有 3′→5′外切核酸酶的校对功能,由 *dnaQ* 编码合成。在体内 ε 亚基的基本功能是控制复制的忠实性,这可由 *dnaQ* 突变体的表型所证实,在细菌株中突变发生的概率增加了 10^3 倍。ε 亚基常与 α 亚基形成一个紧密的 1∶1 复合物,协同发挥功能。在细胞中,大约有 40 个分子的核心酶,其中只有一半被组装到全酶中。核心酶的活性较低,以大约每秒 20 个核苷酸的速度合成 DNA。

(2) β 二聚体 β 亚基是以二聚体的形式存在的。X 射线衍射表明,β 二聚体是环状的,在其间有一个足以容纳 DNA 双螺旋的大空间(图 4-10)。β 二聚体环绕着 DNA,并能在 DNA 上自由滑动,构成一个滑动钳。滑动钳可把全酶束缚在模板 DNA 上,从而保证高度的进行性(processivity,即全酶结合于单链模板不解离的能力)。当 PolIII 结合 β 钳时,全酶的进行性可从每次聚合 10 个核苷酸提高到每次聚合 50 000 个核苷酸,聚合反应速度从每秒 20 个核苷酸上升到每秒 750 个核苷酸。因此 β 钳的存在对大肠杆菌染色体的高效复制是十分重要的。β 亚基可以很容易地从全酶中脱离。与核心酶相比,在细胞中 β 亚基的含量很丰富,大约每个细胞中含 300 个二聚体。

(3) γ 复合物 γ 复合物(γ complex)是 β 滑动钳的载体(clamp loader),可帮助 β 滑动钳结合到 DNA 上。γ 复合物含有 5 个不同的亚基,形成一种化学计量为 $\gamma_2\delta_1\delta'_1\chi_1\psi_1$ 的结构。δ 和 δ′亚基是由不同基因编码产生的不同蛋白质。γ 复合物有依赖 DNA 的 ATP 酶活性。β 二聚体自己并不能组装到 DNA 上,它是通过 γ 复合物与 ATP 协同作用催化 ATP 的水解而组装到 DNA 上的。

(4) τ 二聚体(τ dimmer) 连接蛋白(τ_2)可结合两个核心酶分子和一个 γ 复合物。研究表明,τ 与 γ 是由相同的基因编码产生的,但在翻译时框架有所改变,所以 γ 与 τ 在数量上是相等的。τ 亚基是一个依赖于 DNA 的 ATP 酶,τ 二聚体可结合两个核心酶,每个 τ 亚基结合一个核心酶,形成 PolIII′亚聚集体。在一个分子结构中存在两个聚合酶表明复制性的(replicative)聚合酶成对起作用,协同复制双螺旋 DNA 的两条链。τ 亚基也结合 γ 复合物,PolIII* 是 DNA 聚合酶 III 稳定的亚聚集体,它由两个核心酶,一个 τ 二聚体和一个 γ 复合物构成。PolIII* 包含 9 个不同的亚基,它仅仅缺少 β 亚基。由于缺少 β 亚基,

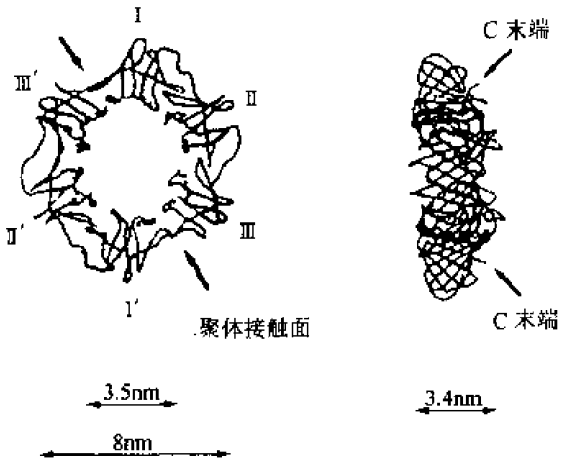


图 4-10 β二聚体结构模型

PolIII* 与全酶相比进行性明显降低，在 PolIII* 上加入 β 亚基可重新产生全酶并恢复其进行性。

人们曾经认为核心酶是由 θ 亚基二聚化的，τ 二聚体结合一个核心酶而 γ 复合物结合另一个核心酶。证据是 τ 和 γ 竞争性地结合包被了 SSB 蛋白的 ssDNA 上的核心酶和 β 亚基。但现在的研究表明核心酶是以单体形式存在。τ 和 γ 的竞争可能是它们都与包被 SSB 蛋白的 ssDNA 结合的缘故。

γ 复合物是一个不对称的结构，因为其中有 4 个亚基是单拷贝，它们相对两个核心聚合酶不对称分布使全酶具有不对称结构，导致在全酶上两个核心聚合酶功能上的不对称性。利用 ATP 的类似物——ATPγS，可证明这种不对称性。与 ATP 不同，当存在 ATPγS 时只有一半全酶可结合引发模板；而当全酶利用 ATP 结合 DNA 后，用 ATPγS 处理可导致一半酶脱落。这表明全酶中的两个聚合酶是不同的，只有一个可利用 ATPγS。两个聚合酶功能上的不对称性将使 DNA 两条链合成表现出不同的性质。

由上可见，DNA 聚合酶 III 全酶的功能是复制大肠杆菌的染色体。DNA 聚合酶 III 与 DNA 聚合酶 I 有很多不同之处；它具有多亚基的结构，聚合反应需要 ATP 的存在，反应有很高的速度和高度的进行性。PolIII 全酶的 10 个亚基协同发挥功能，并与其他同复制有关的蛋白质一起在 30～40 分钟内完成大肠杆菌全长 4.4 Mb 染色体的复制。

2. DNA 复制的忠实性

复制过程的忠实性是保证生物体遗传信息准确传递的一个必要条件，它取决于碱基配对的专一性。每次复制碱基对发生错误的概率是 10^{-3}，而在细菌中实际的错配率为 10^{-8}～10^{-10}，这意味着每个基因组经过 1000 次细菌的复制循环只发生一个错误或每代遗传的错误率只有 10^{-6}。

DNA 聚合酶可通过下述两种方式提高互补碱基选择的专一性。第一，通过专一性识别使即将加入的碱基与模板上的碱基严格互补，这种方式用来控制合成前的错误；第二，当发现存在着错配时，可切除新加入的碱基，这种方式叫校对控制。两种方式可单独起作用，也可共同起作用。由此可见，复制过程中碱基配对受双重核对作用控制；聚合酶的选

择作用和 3′-5′外切核酸酶的校正(proofreading)作用。

所有细菌的聚合酶都有与 DNA 合成方向相反的 3′→5′外切酶活性,该活性提供了如图 4-11 所示的校对功能。在链延伸阶段,一个前体核苷酸进入生长链的末端,二酯键形成后,聚合酶自动向前移,为下一个前体核苷酸的进入做准备。如果此时出现错误,聚合酶就会返回并切除最后一个加入的碱基,使正确的前体核苷酸进入该位置。

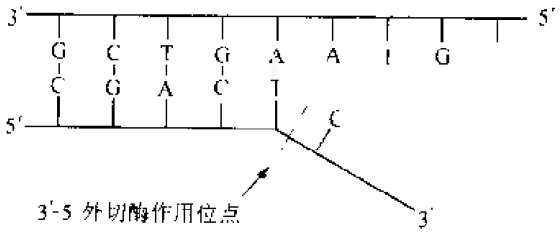


图 4-11 3′→5′外切酶活性的校对功能

不同的 DNA 聚合酶以不同方式处理聚合与校对之间的关系。在一些酶中,两种活性处于同一蛋白质亚基中的不同位置上;而在另一些酶中,两种活性处于同一蛋白质的不同亚基上。实际上,错误碱基的去除要比图 4-11 复杂的多,因为催化切除反应与催化聚合反应的活性处于蛋白质的不同部位。

DNA 聚合酶在复制过程中造成的错误除了不正确配对造成的核苷酸取代外,还包括由于插入或缺失额外的核苷酸造成的框架移位。DNA 聚合酶 3′→5′校正活性也可除去造成框架移位的额外碱基。产生框架移位错误的概率还受聚合酶进行性的影响,这对同聚序列(homopolymeric stretch)的复制尤其重要。例如,当复制一个 $dT_n:dA_n$ 的长序列时,复制滑动可改变同聚序列的长度。通常提高 DNA 聚合酶的进行性可降低框架移位错误发生的可能性。多亚基 DNA 聚合酶进行性的提高是由特殊亚基决定的,而该亚基对于酶的催化活性并不是必需的。上面提到的 DNA 聚合酶 III 的进行性因子是它的 β 亚基,细菌噬菌体 T4 和真核复制酶也有各自的进行性因子,分别为 gp45 蛋白和增殖性细胞核抗原 PCNA,这些进行性因子也有与 β 二聚体类似的可在 DNA 上滑动的环状结构。

3. 其他生物的聚合酶

一些噬菌体也编码 DNA 聚合酶,它们是 T4、T5、T7、SPO1。这些酶都有 5′→3′合成活性和 3′→5′外切校正活性。噬菌体聚合酶多肽与其他蛋白质在噬菌体或宿主的复制起点相结合形成完整活性的酶,在噬菌体中编码聚合酶的基因发生突变会阻止噬菌体的发育。

目前,在真核生物中发现了 5 种不同的 DNA 聚合酶,分别为 α、β、γ、δ、ε 所示。α、β、δ、ε 位于细胞核中,而 γ 是线粒体酶。其中聚合酶 α、δ、ε 同属一类 DNA 聚合酶,有共同的结构和催化特征。

4. DNA 聚合酶的引物

DNA 聚合酶的一个共同特征是它们只能催化脱氧核糖核苷酸加到已有核酸链的游离 3′-羟基上,而不能从游离核苷酸起始 DNA 链的合成。所以,它们需要引物(primer)来

提供 3′-OH 末端，然后在其上加入核苷酸来延伸 DNA 链。引物有多种形式，通常细胞 DNA 和一些病毒先在模板上合成一段 RNA 序列，然后通过 DNA 聚合酶延伸 RNA 链的 3′ OH。反转录病毒也利用这种机制引发 RNA 的反转录过程。而环状双螺旋 DNA 通常在内部形成一个缺口产生引物末端，如 DNA 滚环复制就是这样起始的。这个过程与图 4-8 中的切口移位不同，前者用新合成的链置换已有的链，而后者 DNA 聚合酶从缺口开始合成并降解原有 DNA。此外，一些特殊的蛋白质可以把一个与其相连的核苷酸引入到 DNA 聚合酶的催化位点来直接引发反应。有一些病毒利用这个机制。

5. DNA 聚合酶催化的分子机制

多核苷酸聚合酶（包括 DNA 聚合酶和 RNA 聚合酶）可能是生物进化过程中最早出现的酶，生物体利用它们来准确复制自身的基因组。聚合酶的作用机制必然有以下特征：首先，聚合机制必须是简单且普遍适用；第二，这些酶必须有高度的忠实性，可把基因组的序列信息准确地复制给子代，否则亲代所含的遗传信息将不复存在；最后，复制性的 DNA 聚合酶必须是高度进行性的，可以一次复制整个基因组而不从模板上脱落。人们通过研究，已对 DNA 聚合酶的催化机制及其复制忠实性和进行性的机制有了深刻的了解。由聚合酶与底物复合物晶体的 X 图谱可初步了解聚合酶的结构是如何促使 dNTP 加到引物 模板复合物上的。

PolI 家族的聚合酶有相同的结构特征，具有 N 末端 3′→5′校正作用的外切核酸酶区域和 C 末端聚合酶区域。其聚合酶区域的结构由三个亚区域组成，使人联想到一个半张开的右手，这三个亚区域被称为手指、手掌、拇指，它们形成一个可结合 DNA 的大缝隙，通向聚合酶活性部位。来源于病毒、原核生物、真核生物的聚合酶、反转录酶甚至 RNA 聚合酶都有这样的基本结构。手掌亚区域在缝隙的底部，含有保守的氨基酸序列微区 A、C（存在于所有的核酸聚合酶中）和 B（只存在于依赖 DNA 的聚合酶中），它们构成了活性位点。这些保守氨基酸包括结合参与催化的金属离子的关键酸性氨基酸，还包括与引物末端以及 dNTP 的 α 磷酸相互作用的氨基酸残基。聚合酶缝隙的壁部由手指和拇指亚区域组成。尽管与手掌相比其氨基酸保守性要小，这些亚区域在不同聚合酶中的功能却十分相似。通过晶体结构和模型研究，手指区域可与未复制的单链模板相互作用，同时手指区域的一部分也参与结合进入的底物 dNTP。而拇指亚区域可与模板-引物 DNA 双螺旋相互作用。聚合酶通过保守残基与引物模板复合物相互作用使引物的 3′末端定位在手指和手掌的会合点，并与 dNTP 相对。

人们对嗜热脂肪芽孢杆菌（*B. stearothermophilus*）DNA PolI 进行了研究，通过考察引物-模板和活性部位的结构，揭示聚合酶与 DNA 是如何以一种不依赖序列的方式相互作用的。聚合酶与引物-模板复合物 3′端 4 个碱基对的小沟 DNA 发生了广泛的相互作用。小沟与大沟不同，含有不依赖于核苷酸序列的氢键供体和受体，所以通过小沟接触允许聚合酶结合任何序列。聚合酶结合的 DNA 主要是 B 型结构，但靠近引物末端的 DNA 通常是 A 型结构，该结构具有较低的螺旋扭曲和更宽的小沟，使进入小沟更加容易。

模板和正进入的核苷酸之间新形成的碱基对并不与聚合酶蛋白发生氢键作用，正确的核苷酸专一性是由这个碱基对、聚合酶以及引物末端三者在空间上配合的紧密程度决定的。引入的碱基和模板上与之配对的碱基恰好进入一个仅能容纳 Watson-Crick 配对

的口袋中，口袋里有限的空间限制了 dNTP 的运动，促使其与模板形成标准的 Watson-Crick 碱基对。

聚合酶活性位点上的许多保守残基还对 dNTP 的进入与结合产生了空间限制和静电限制。在 dNTP 的 β 和 γ 磷酸基团上，不与金属作用的氧残基都与聚合酶螺旋上的一个或多个保守残基相互作用。另外，DNA 聚合酶上的一些疏水残基在核糖的 C2′位置构成一个疏水性口袋，可将错误进入的核糖核苷酸驱除。以上结构为 DNA 聚合酶区分掺入的核苷酸种类提供了结构基础。

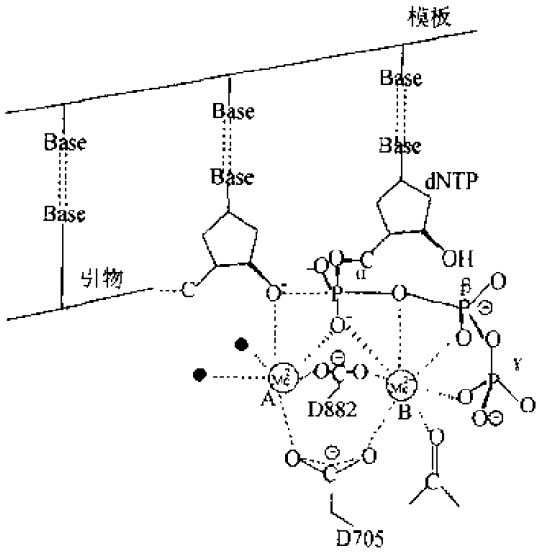


图 4－12　聚合反应的双金属离子假说

研究还发现，dNTP 的进入伴随着两个 Mg 离子的作用。由此提出了核苷酸加入（聚合）反应的双金属离子机制（two metal-ion mechanism）。（图 4－12）聚合酶活性部位位于手掌上，具有 3 个保守的酸性残基，其中两个保守羧基（ASP）可与两个金属离子（Mg^{2+}）结合并使它们固定在活性部位上。这两个 ASP 残基在 DNA 和 RNA 聚合酶中广泛存在。同时，这些金属离子又与 dNTP 的磷酸基团相结合，通过与三个磷酸基团上非酯化的氧相互作用，从而限制了核苷酸的定位。此外，金属离子 A 降低了引物 3′OH 对 H 的亲和力，促进了 3′O^- 亲核结构进攻核苷酸的 α 磷酸基团，从而催化了磷酰基转移反应，最终导致了核苷酸的掺入。而金属离子 B 促进了焦磷酸的释放。这两个离子还通过抵销过渡态产生的负电荷，稳定了五价过渡态结构。除了手掌亚区域的酸性氨基酸残基外，手指亚区域的一些残基也直接参与 PolI 家族聚合酶的催化反应。例如，在嗜热脂肪芽孢杆菌的 DNA PolI 中，酪氨酸残基在构建活性位点的立体结构时发挥了重要的作用，从而保证了正确的碱基配对。

PolI 家族成员的晶体结构显示出在它们的聚合酶区域存在相似的折叠结构，它们包括大肠杆菌 DNA 聚合酶 I 的 Klenow 片段，*Taq* DNA 聚合酶的类似片段，和嗜热脂肪芽孢杆菌 DNA 聚合酶。另一种非同源的 DNA 聚合酶——哺乳动物的 DNA 聚合酶 β 也具有类似的机制。在该酶中发现了相似的引物-模板 DNA、dNTP 和两个 Mg^{2+} 的排列。在 Pol α 家族 DNA 聚合酶、T7 RNA 聚合酶、HIV 病毒反转录酶和脊髓灰质炎病毒的依赖

RNA 的 RNA 聚合酶中也都存在着相似的催化区域结构、ASP 残基和金属离子连接方式。这表明以上这些多核苷酸聚合酶可能采取相同的双金属离子机制进行催化。其他磷酸转移反应，如聚合酶上起校正作用的外切核酸酶，也可能采用类似的双金属机制进行催化。该机制也适用于由 RNA 形成的 RNA 和 DNA 聚合酶，因此在远古的 RNA 世界里也可能发挥重要作用。

DNA 聚合酶与引物模板复合物结构的确定也部分解释了聚合酶复制 DNA 的忠实性。核苷酸的掺入是受模板指导的，所以 DNA 聚合酶在每个催化循环中都要改变它的核苷酸底物专一性，该过程的错误概率仅有 10^{-5}～10^{-6}。这个结构模型描述了分子识别的细节：聚合酶的氨基酸残基可与 DNA 小沟中的氢键受体相互作用，4 种 Watson-Crick 碱基对的相同位置都有氢键受体存在，所以这种相互作用与所复制的 DNA 的碱基序列无关。但只有形成正确的 Watson-Crick 碱基配对时，这种相互作用才会发生。如果缺少这种必要的相互作用，含有错配碱基的引物模板复合物对聚合酶的结合能力会下降，或者不能正确定位于 DNA 聚合酶活性位点，导致 DNA 合成的速度明显放慢，从而触发 DNA 聚合酶从 DNA 合成到校正的转变。错误掺入也促使引物-模板双螺旋的熔解，产生一个具有 4～5 nt 的 3′单链末端的 DNA 分子，这种 DNA 结构是外切核酸酶的适宜底物。大肠杆菌 PolI 的聚合酶和外切核酸酶活性位点距离 3nm，但通过一个共同的 DNA 结合缝隙相连接。这种安排表明生长链的 3′末端必须从聚合酶活性位点转移到校正的活性位点以切除错配。

三、DNA 连接酶

DNA 聚合酶只能催化多核苷酸链的延长反应，不能使链与链之间连接。因此，在复制过程中，必然存在一种酶，能催化链的两个末端之间形成共价连接，这个酶就是 DNA 连接酶(DNA ligase)。DNA 连接酶能催化双链 DNA 切口处的 5′-磷酸基和 3′-羟基生成磷酸二酯键。这种反应需要供给能量，大肠杆菌和其他细菌的 DNA 连接酶以 NAD 作为能量来源，动物细胞和噬菌体的连接酶则以 ATP 作为能量来源。连接反应包括两步：首先形成一个酶- AMP 复合物，这个复合物通过 AMP 与缺口处的 5′磷酸相连，然后 5′磷酸与缺口的 3′羟基末端形成磷酸二酯键，并放出酶和 AMP。

四、螺　旋　酶

在体外，大多数 DNA 以双螺旋的形式存在。然而，在 DNA 复制、修复、重组及接合过程中，DNA 两条互补链必须部分分开以形成单链 DNA 中间体，即解旋。该过程是由 DNA 螺旋酶催化完成的。螺旋酶(helicase) 又称为解旋酶，是促进 DNA 的两条互补链分离的酶。DNA 螺旋酶是 DNA 复制、重组、修复过程中关键的发动蛋白。螺旋酶具有依赖 DNA 的 ATPase 活性，可把 ATP 水解产生的自由能转化为解旋 DNA 并使螺旋酶本身沿 DNA 链移动的机械能。在解旋过程中，这些酶打断互补碱基配对的氢键并结合、水解 ATP，以 500～1000 bp/秒 的速率沿 DNA 链解旋 DNA 双螺旋。

在原核和真核生物、细菌噬菌体和病毒中都能检测到 DNA 螺旋酶的存在。大多数

有机体的细胞能编码多种螺旋酶，这些螺旋酶的功能相互重叠或平行。例如，大肠杆菌能编码至少12种螺旋酶，酵母至少6种。DNA螺旋酶也是真核生物转录复合物的组分，还有一些螺旋酶具有RNA螺旋酶活性，在真核生物的转录起始中发挥作用。有些螺旋酶还能使RNA/DNA双螺旋解旋。各种螺旋酶在DNA解旋和沿DNA链移动方面具有一些共同的机制。其主要的特征是DNA螺旋酶都以寡聚体的形式发挥功能，通常是二聚体或六聚体(表4-5)。每个亚基具有一个潜在的DNA结合位点，从而每分子螺旋酶具有多个DNA结合位点，这是螺旋酶实现催化功能的关键。在聚合状态已被详细研究的螺旋酶中，除了RecBCD酶，其他螺旋酶多以寡聚体的形式存在。大肠杆菌DnaB、大肠杆菌RuvB、噬菌体T7基因*4*螺旋酶/引发酶、噬菌体T4基因*41*螺旋酶、SV40 T抗原和大肠杆菌Rho蛋白是六聚体，而大肠杆菌的Rep蛋白是一个二聚体。DNA解旋和转位的分子机制要求螺旋酶至少应具有两个DNA结合位点。

复制叉是由复制性的DNA螺旋酶的作用产生的，尽管RNA聚合酶也可熔解DNA双螺旋。只有少数的螺旋酶专一性地参与复制叉的产生。最近对不同生物复制性的螺旋酶如大肠杆菌的DnaB、SV40的T抗原和噬菌体T4和T7螺旋酶的研究表明，尽管这4种蛋白质之间没有明显的序列相似性，但它们都有共同六聚体结构而且生物化学性质也相似，如以高的进行性解旋以及与相应的复制性DNA聚合酶协同作用。这些性质是产生复制叉的螺旋酶所普遍具有的。

表4-5　螺旋酶的寡聚性质

螺旋酶	聚集情况	螺旋酶	聚集情况
SV40大T抗原	六聚体	T4噬菌体基因*41*蛋白	六聚体
T7噬菌体基因*4*蛋白	六聚体	大肠杆菌RuvB	六聚体
大肠杆菌Rho	六聚体	大肠杆菌螺旋酶II	寡聚体
大肠杆菌TraI	寡聚体	大肠杆菌RecBCD	异三聚体或六聚体
大肠杆菌RecB	寡聚体(二聚体或四聚体)	大肠杆菌Rep	二聚体
大肠杆菌螺旋酶II(UvrD)	二聚体或大的寡聚体	大肠杆菌Rep/UvrD	异二聚体
大肠杆菌UvrAB	A_2B异三聚体	HeLa螺旋酶	二聚体
HSV-1起点识别蛋白	二聚体	HSV-1螺旋酶/引发酶	异二聚体

电镜图象重建技术显示六聚体复制性的螺旋酶可产生围绕DNA的蛋白环。例如噬菌体T7的螺旋酶是一个具有两个不同表面的六聚体环，其外径13nm，具有2.5～3.0nm的洞，可覆盖30nt的单链DNA。在大肠杆菌DnaB、T4螺旋酶、SV40的T抗原-DNA复合物中也有类似的结构。但不清楚这些螺旋酶可包围一条DNA链还是可包围DNA双链，可能在不同的螺旋酶中具体情况不同。这些螺旋酶围绕DNA的事实解释它们在复制叉上近似无限的进行性。因此，六聚体的DNA螺旋酶与DNA聚合酶的滑动钳一样，都能通过与DNA拓扑连接实现高度的进行性。一旦这些螺旋酶结合DNA，双螺旋将连续熔解直到一些活化过程终止螺旋酶活性，在细菌中序列专一性的终止蛋白提供了这种螺旋酶路障。

螺旋酶催化的DNA解旋机制可分为两大类：主动机制和被动机制。在主动机制中螺旋酶使双螺旋DNA去稳定，而被动机制是通过结合双链DNA上暂时形成ss-DNA间接促进解旋的。被动机制中螺旋酶只有一个结合位点与ss-DNA结合，结合后沿ss-DNA向双螺旋运动使未解旋的双链分开。主动机制中螺旋酶至少要有两个结合位点，该机制

又分为两亚类:第一种主动机制是螺旋酶直接与 dsDNA 结合,并使部分碱基去稳定。该过程包括 ATP 的结合、水解、产物释放,从而引起螺旋酶的结构变化,实现解旋。该机制需要酶在催化的中间过程中能同时结合 ss 和 dsDNA,它又包括蠕虫机制和滚动机制两种类型。第二种主动机制是扭曲机制,螺旋酶不直接与双螺旋 DNA 结合,而是在 ss /ds DNA 连接处同时结合两条单链,通过 ATP 诱导结构变化而破坏邻近的双链区域使之解旋。对大肠杆菌 Rep 同源二聚体螺旋酶的解旋机制研究得最彻底,它是大肠杆菌中最主要的螺旋酶,通过 Rep 二聚体以主动的滚动机制进行 DNA 解旋和移位。滚动机制适用于所有同聚螺旋酶;它需要螺旋酶有两个相同的 DNA 结合位点,每个位点必须既能结合 ssDNA 又能结合 dsDNA;并假设 Rep 二聚体呈 C2 对称,每个亚基以相对磷酸骨架特定的极性结合 ssDNA。在滚动机制中,始终有一个亚基结合在复制叉处 3′方向的 ss DNA上,另一个亚基根据不同情况可结合在同一条单链上或复制叉前的双链区域。每个亚基可交替地结合 ssDNA 和 dsDNA,这是由 ATP 和 ADP 结合造成空间结构的变化来控制的。具体解旋过程如图 4 13 所示:中间体 I 二聚体的两个亚基都结合到 3′ssDNA 上,形成 P_2S_2 复合物;与 ATP 结合后,其中一个亚基对 ssDNA 的亲和性下降,对双螺旋 DNA 的亲和性提高,通过中间体 II 产生中间体 III(P_2SD);在 P_2SD 复合物中,Rep 二聚体同时结合在复制叉前的双螺旋和 3′ssDNA 链上;然后通过 P_2SD 复合物水解 ATP 导致蛋白质结构发生变化,在双螺旋中产生多个碱基的去稳定,置换出 5′单链;同时 Rep 亚基还结合在 3′单链上,从而产生了 P_2S_2 复合物;随后的 ADP 和无机磷酸的释放完成了产生中间体 I′的循环。因为 Rep 是二聚体,所以 I 与 I′在功能上是一致的。

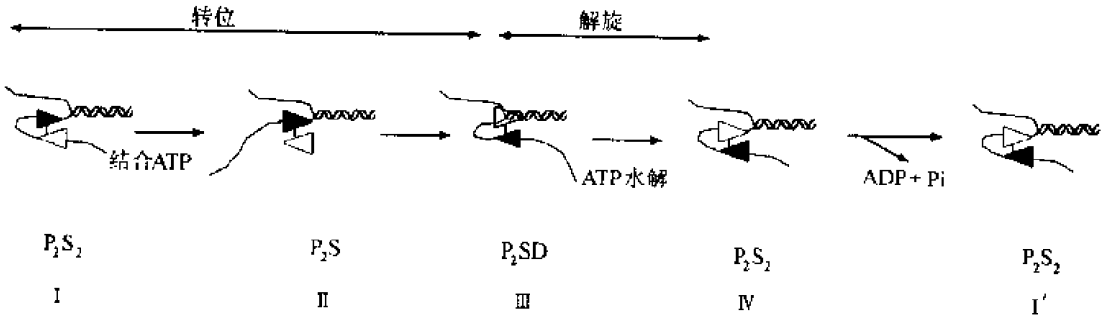


图 4-13 螺旋酶催化的 DNA 解旋机制

在这个模型中,通过 Rep 二聚体的翻转及与 ATP 结合、水解的循环,完成了双螺旋 DNA 的解旋。每次催化循环中有多个碱基对被解旋,而且有功能的二聚体至少有一个亚基与 3′(5′)DNA 结合,这个机制可保持解旋的高进行性。滚动机制不同于蠕虫机制,后者也需要螺旋酶有两个 DNA 结合位点,但这两个结合位点是不同的,与 ssDNA 的结合是有方向性的,这使螺旋酶具有了极性。如头位点可交替与 ssDNA 和双螺旋 DNA 结合,而尾位点只与 ssDNA 结合,以蠕动机制解链。

五、DNA 拓扑异构酶

复制过程中,DNA 双螺旋的解旋使复制叉前面产生正超螺旋,如果不以某种方式释

放这种应力，将妨碍复制叉前进；在大肠杆菌中，如果这些超螺旋传递到分子末端才被释放，DNA 分子每分钟必须转 20 000 转左右，而且环状分子没有自由末端。拓扑异构酶可以使超螺旋分子松弛，从而解决上述问题。

在大肠杆菌中，起主要作用的拓扑异构酶是拓扑异构酶 II，又被称为 DNA 旋转酶（DNA gyrase）。它由两个 A 亚基和两个 B 亚基组成，B 亚基可以结合 ATP，具有 ATP 酶活性。拓扑异构酶 II 能产生负超螺旋以抵销复制叉移动所产生的正超螺旋，它每次能使 DNA 双链同时断裂，然后使断裂末端与两个 A 亚基共价连接，而不至于造成 DNA 分子的开环，再让另一条双链 DNA 穿过缺口，将已切断的末端重新连接起来。每次催化循环引起的超螺旋数目变化为 2。通过控制正超螺旋产生及其转化为负超螺旋的速率平衡，旋转酶可防止正超螺旋的积累，使 DNA 片段以负超螺旋形式存在。大肠杆菌旋转酶温度敏感型菌株在非许可温度下，DNA 合成速度立即下降。此外，在环状 DNA 复制的最后一步，旋转酶可分开复制后的两个环状子代 DNA 链。

细菌中的其他拓扑异构酶也可能在复制过程中起作用。除细菌拓扑异构酶 I 外，4 个已知的拓扑异构酶中有 3 个支持复制中新 DNA 链的延伸。拓扑异构酶 IV 可有效去除两种超螺旋，在高浓度时能防止产生过度的超螺旋。大肠杆菌拓扑异构酶 III 虽不能去除正超螺旋但也能支持亲链解旋。理论上，DNA 拓扑异构酶可在含超螺旋的任何拓扑区域发挥作用，去除复制产生的正超螺旋。例如，θ 复制中间体的正超螺旋可被 I、II 型拓扑异构酶在 θ 的非复制区除去。

在酵母中，拓扑异构酶 I 可能是复制时主要的拓扑异构酶，但它的功能可被拓扑异构酶 II 取代。新链延伸需拓扑异构酶 I、II 的共同参与，拓扑异构酶 III 自身不足以完成任务。

六、单链结合蛋白

在 DNA 链上，前进的解旋酶后存在着单链区域，为避免单链退火或形成链内氢键，单链区必须与单链结合蛋白（single-strand binding protein，SSB）结合。SSB 的另一个生物学功能可能是在复制中与单链 DNA 高亲和性地结合并保护单链 DNA 部分不被核酸酶水解。作为辅助蛋白，SSB 在 DNA 的复制、重组、修复等过程中起着关键的作用。除了能以高亲和力结合 ssDNA 外，SSB 还有其他一些特征：SSB 与 DNA 结合时无序列专一性；SSB 可加强单链 DNA 对 DNA 聚合酶的亲和力；可与其他有关复制的酶如螺旋酶作用，或促进它们发挥自身作用；复制中所需的 SSB 的数量与 ssDNA 模板的数量呈化学计量关系；当 DNA 聚合酶前进时，会置换出结合在单链上的 SSB 蛋白。

最早发现的 SSB 是 T4 噬菌体的基因 *32* 编码的蛋白质，随后在大肠杆菌、其他细菌、接合细菌质粒、其他噬菌体及真核生物中也发现了 SSB 功能上的类似物。

原核细胞中的 SSB 主要有 T4 噬菌体基因 *32* 的编码产物（gp32）、大肠杆菌的 SSB 等。gp32 以单体形式存在，每个细胞含有 10 000 个拷贝。其单体能使复制复合物前面的双链 DNA 去稳定；保护单链 DNA 不被核酸酶降解；防止单链 DNA 模板形成双螺旋发夹结构。gp32 可与 T4 复制中的其他蛋白质相互作用，gp32 的功能不能被大肠杆菌的 SSB 取代。gp32 对 ssDNA 的结合有高度的协同性，可形成连续的蛋白质串，很容易使单链

DNA 饱和。

大肠杆菌中的 SSB 是同源四聚体，其单体有 177 个氨基酸，每个细胞含有 800 个拷贝。SSB 参与引发体的组装；消除使 PolIII 停止前进的二级结构；促进大肠杆菌 PolII、PolIII 的活性，但不能促进 PolI 和 T4 聚合酶的活性；能提高 DNA 复制的忠实性。SSB 与单链核苷酸有多种结合方式。每个四聚体可结合 35±3 或 65±5 个核苷酸。在 $(SSB)_{35}$ 复合物中，单链 DNA 只能与 SSB 的两个亚基作用；在 $(SSB)_{65}$ 复合物中，可与所有 4 个亚基作用。接合质粒和线粒体中的 SSB 与大肠杆菌中的 SSB 相似，而真核细胞中的 SSB 明显不同于大肠杆菌的 SSB 和 gp32。

第三节　复制的起始

所有已知的 DNA 分子都在复制起点处起始一轮复制。复制是在起始阶段进行控制的，复制子增殖依赖于起始过程中在起点处发生的相互作用。一旦复制起始后，它就会进行到底，直到整个基因组都复制。起始的频率由复制起点和调控蛋白的相互作用控制。在一个细胞周期中，每个复制子只启动一次。

起始有两种方式：从头起始（*de novo* initiation），DNA 链的合成从头开始；共价延伸（covalent extension），新链从原有的亲链上开始合成。我们后面要讲到的滚环复制就是通过共价延伸起始的。在滚环复制中，在一条断裂的亲本链的 3′-OH 上不断地进行 DNA 聚合作用。由于断裂的 3′末端仍然与互补链形成双螺旋结构，因而不需要合成 RNA 引物。

DNA 双螺旋复制的从头起始包括一些连续的步骤：

1）DNA 分子上一段短的区域发生熔解反应，双链分离形成单链区域。复制的起始同螺旋酶的解旋和 SSB 蛋白的作用紧密相关，螺旋酶由 ATP 水解所驱动，并可能借助 SSB 蛋白的结合把螺旋解开，单链部分被 SSB 所覆盖。

2）解旋点沿 DNA 移动标志着复制叉的产生，复制叉在 DNA 链延伸时将连续移动。

3）引发反应和 RNA 引物的合成。前者需要多种蛋白质形成的复合物，即预引发体（preprimosome）和引物合成酶参与。在前导链上只需一次引物 RNA 合成，发生在复制起点，但在后随链上必须合成一系列这样的引物。因为每个冈崎片段都有自己的复制起点，所以每个冈崎片段都需一个 RNA 引物来起始 DNA 的合成。引物的长度通常为 10nt，它提供了 DNA 复制酶延伸所需要的 3′羟基末端，DNA 复制酶可在其上进行聚合反应，直到完成冈崎片段的合成。

下面以大肠杆菌的复制起始为例说明原核生物复制起始的情况（图 4-14）。基因组 DNA 的合成包括多种蛋白质的高度协同作用。这些蛋白质以极高的速度装配两条新合成的 DNA 链。在大肠杆菌中大约每秒 1000 核苷酸。

一、预引发体的形成

复制起始需要多种蛋白质的参与，它们构成了预引发体的组分，其中包括专一性识别起点序列的蛋白质（起始蛋白）；参与组装过程的蛋白质，这些蛋白质在随后的 DNA 合成

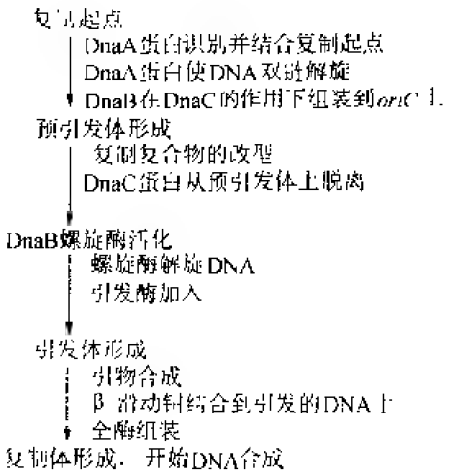


图 4-14　复制起始过程的示意

中往往不发挥作用，如装载(loading)因子和改型(remodeling)因子；以及复制机器的组分，如 DNA 螺旋酶。这些因子共同作用控制复制叉组装的时间和位点，它们还作为细胞循环调控的靶子。起始因子在原核生物、噬菌体、真核生物中功能上是保守的。人们用大肠杆菌的慢停突变型鉴定了许多产物与复制起始有关的基因，并利用含有大肠杆菌 *oriC* 序列的质粒构成的无细胞体系研究在 *oriC* 处的复制起始。这些实验表明：*oriC* 的复制起始需要 DnaA、DnaB、DnaC、HU、gyrase 和 SSB 的参与，它们形成复合物协同作用使双螺旋上出现连续的解旋。

大肠杆菌的起始蛋白是 DnaA 蛋白，它可识别复制起点并与之结合，这种结合是随后的起始事件的基础。一旦结合 DNA，起始蛋白会获得额外的性质：①它可促进 DNA 解旋或使邻近的 DNA 发生扭曲使 DNA 螺旋酶能够进入；②它们可帮助与复制叉组装和功能有关的其他因子进入。

大肠杆菌起始复合物的形成起始于 DnaA 蛋白在 *oriC* 上的结合，大肠杆菌的双向复制起点 *oriC* 包括 245 bp 的序列，由 9 bp 和 13 bp 重复序列组成。起点经过如图 4-15 的一系列活化步骤，首先是 DnaA 的结合。4 个 9 bp 的一致序列位于 *oriC* 位点的右侧，提供了 DnaA 的起始结合位点，DnaA 之间协同作用形成一个由 20～40 个单体构成的核心，*oriC* DNA 缠绕在这个核心上。尽管起始因子结合起点对复制起始是关键的，但这种结合并不足以引发复制起始。首先起始因子的 DNA 结合位点并不是实际的起点，其次当 DnaA 结合起点时，起点是不活化的。因此，起始蛋白结合起点并不是复制起始过程的决定步骤。

复制叉组装的第二个关键步骤是 DNA 双链解旋，以装载复制机器并提供 DNA 合成的模板。DnaA 蛋白作用于 *oriC* 左侧的 3 个 13 bp 的串联重复序列，使 3 个 13 bp 重复序列处的 DNA 链熔解形成一个开放的复合物，以促使反应进一步进行，以上两种重复序列都富含 A、T，这是起点发挥功能所需要的。DnaA 是 ATP 结合蛋白，ATP 对 DNA 结合并不是必须的，因为 ADP-DnaA 蛋白也可以形成起始复合物，但不能打开 *oriC* 处的 DNA 双螺旋。DnaA 还能与单链区域的一条链发生专一性的相互作用，因此除了专一性地识别双链 DNA 结合位点外，起始蛋白在起点解旋过程中还可以作为单链 DNA 结合蛋白发

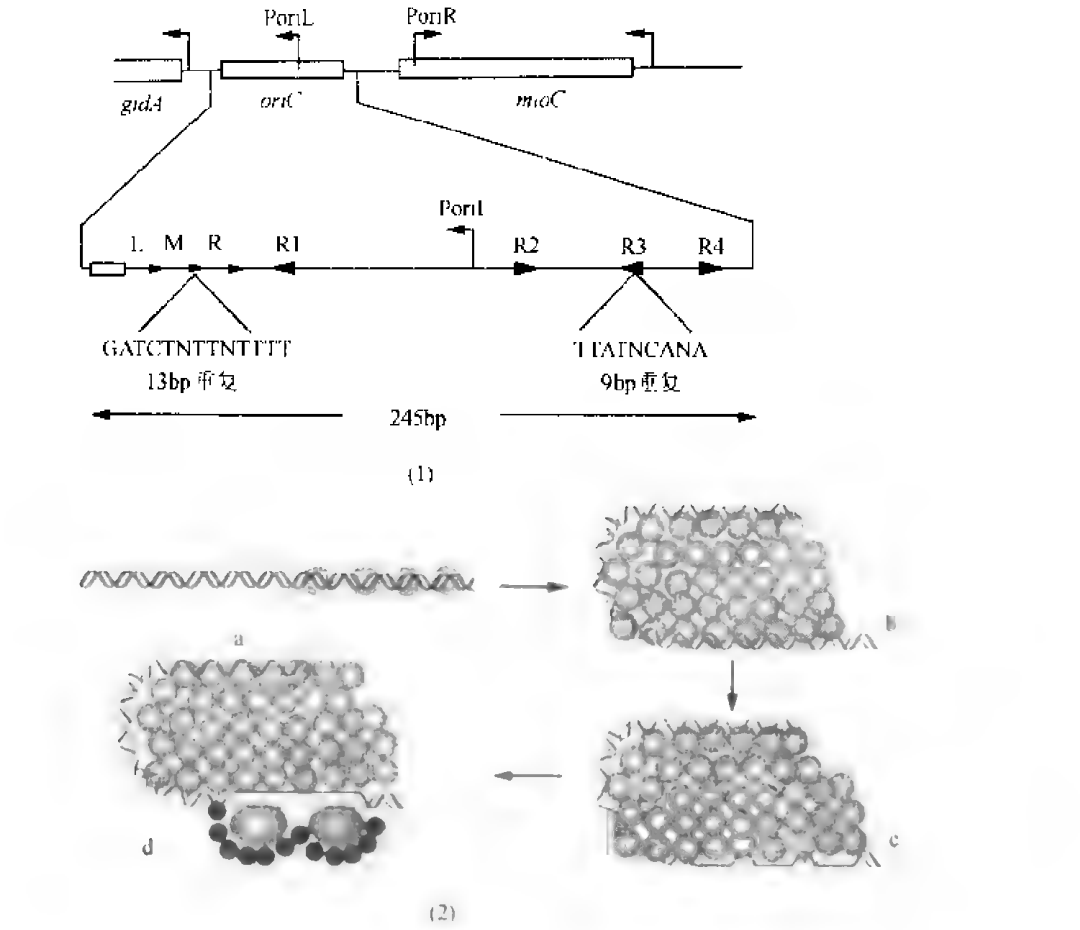


图 4-15 大肠杆菌复制的预引发反应

(1)大肠杆菌的复制起点 R1～R4 是 9 bp 重复序列，为 DnaA 结合位点。L、M、R 是 13 bp 重复序列。上端的箭头代表转录的方向。(2)预引发过程及起始蛋白。a. DnaA 单体结合在 9 bp 重复序列上；b. 20～40 个 DnaA 单体形成大的聚集体；c. 在 13 bp 处导致 DNA 链分离；d. DnaB/DnaC 加入，使 DNA 链进一步解旋。

挥作用。

一旦结合起点 DNA，起始蛋白将促使参与复制叉组装的其他因子的结合，这个过程的主要目的是使 DnaB 进入复制叉的组装位点。DnaB 实际上是使 DNA 解链的螺旋酶，它能识别复制叉处潜在的单链结构，它通过取代 13 bp 重复序列处的 DnaA，并与 DNA 旋转酶协同作用，促使大范围的解链，同时 SSB 蛋白稳定形成的单链 DNA。DnaB 是一个六聚体蛋白质，与 6 个 DnaC 蛋白结合形成 DnaB/DnaC 复合体，这是一个分子量 480 kDa、直径 6 nm 的球状蛋白聚集体。这个复合物与 *oriC* 结合形成预引发体(preprimosome)，从而形成了一个双向复制叉。其中 DnaC 是 DnaB 的装载因子，伴随 DnaB 六聚体结合复制叉组装位点，此外它还使螺旋酶保持无活性的状态。因此，起始因子与单链 DNA 的相互作用，以及 DnaB/DnaC 复合物与起始蛋白 DnaA 的蛋白质-蛋白质相互作用对复制叉的组装都是必要的。

二、复制复合物的改型

起始因子在DNA复制起点的组装需要起点结合复合物保持高度的稳定性，但是DNA复制却需要起始蛋白-DNA复合物的组分保持自由状态以完成它们的功能。因此，在复制起始过程中，需要这些复合物改型或去组装以使稳定的起点结合复合物转化为活动的复制机器。在这个转化中，使起始因子组装为起点复合物所必需的蛋白质-蛋白质和蛋白质-DNA之间的相互作用将被破坏。在大肠杆菌的DNA复制中，通过改变起始复合物各组分的ATP结合状态来改变其组成形式，例如，大肠杆菌的DnaC蛋白在*oriC*必须结合ATP才能与DnaB相互作用并装载DnaB，当ATP水解或释放时，DnaC会离开复合物。

起始过程的最主要的改型步骤导致了复制叉处螺旋酶的活化。螺旋酶装载的机制还不清楚，但因为复制性的DNA螺旋酶可能围绕DNA的一条链或双链，可能需要开环过程。螺旋酶活化可能是装载复制叉其他组分的关键步骤，一旦螺旋酶活化，它将连续熔解双螺旋DNA。

三、引发体的形成

高度解链的模板与蛋白质的复合体促进引发酶加入进来形成引发体(primosome)，合成引物RNA。

在细菌中，已知有两种合成RNA的酶，一种是RNA聚合酶，这是转录时所使用的酶。另一种是引发酶(primase，是*dnaG*基因的产物)，它是一个单亚基的多肽。这两种酶可以通过对抗生素利福霉素的敏感性不同而区分：RNA聚合酶对利福霉素敏感，而引发酶则不敏感。前导链的合成可被利福霉素抑制，而后随链中冈崎片段的起始不被抑制，表明后随链引物的合成是由引发酶进行的。

引发酶不同于参与DNA复制的聚合酶，它能重新起始一个新核苷酸链的合成。引发酶在模板DNA的特定位点起始链合成，这种序列的偏爱使引发酶与复制性的DNA聚合酶区别开。这些起始位点对应于简并的(degenerate)二核苷酸序列，因此有许多位点可以起始引物的合成。不同的引发酶识别不同的序列，一些引发酶中有锌指样的DNA结合区域参与DNA序列的选择。大多数引发酶可以使用核糖核苷酸和脱氧核糖核苷酸作为底物，但是因为细胞中有大量的核糖核苷酸，所以引物通常是RNA。引发酶的进行性很有限，通常合成的链短于12 nt。

复制的起始也需要RNA聚合酶的参与，由其进行的转录在*oriC*起点附近终止并产生了一段RNA，这段RNA可能作为引物，提供引发DNA聚合酶III反应的3′羟基；也可能通过转录步骤导致起点处DNA结构发生变化，以利于引发体的形成，这种情况称为转录激活。在第二种情况下，前导链的引物合成可能也是由引发酶来完成的(图4-16)。

DnaG引发酶与预引发复合物中的DnaB螺旋酶结合形成可移动的复合物，并合成RNA引物。DnaB螺旋酶和引发酶之间的蛋白质接触对复制叉功能是重要的，引发酶的聚合酶活性通过与DnaB相互作用而激活。与DnaB和DNA PolIII全酶不同，引发酶并

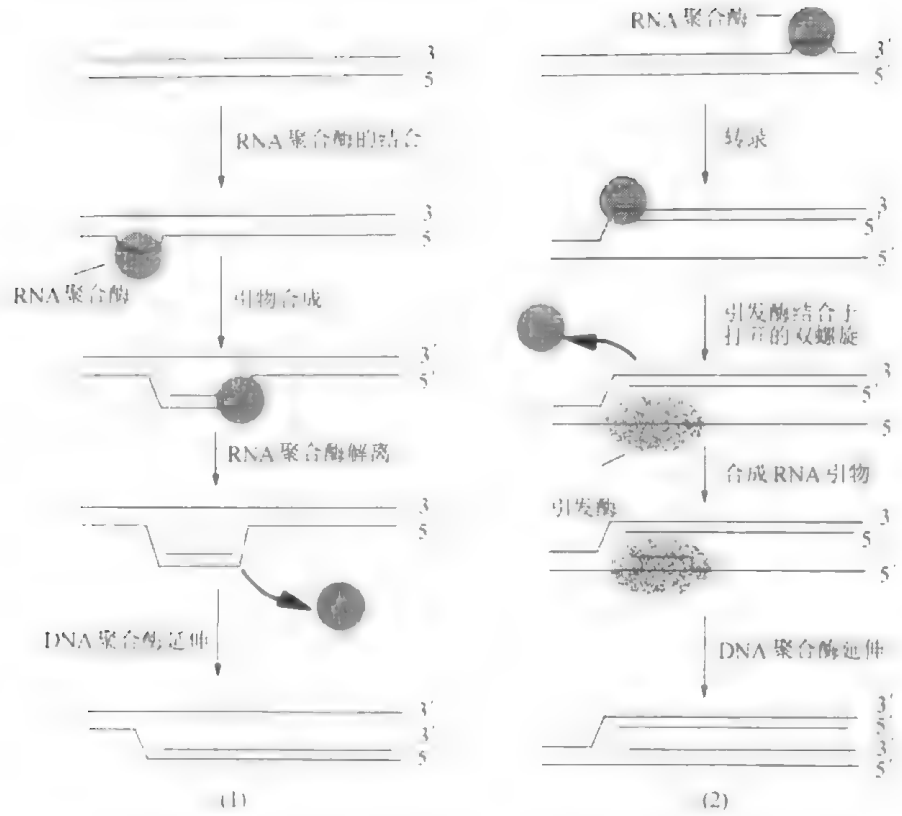


图 4-16 RNA 聚合酶在复制中的作用

(1)RNA 聚合酶转录出 RNA 短链作为引物;(2) RNA 聚合酶转录出的 RNA 短链分开了 DNA 双螺旋,使引发体得以形成。然后由引发酶合成引物。

不作为复制叉蛋白质复合物的稳定组分移动,在每个引发事件中它从溶液中重新组装。影响引发酶- DnaB 相互作用的突变可改变冈崎片段的长度,因此引发酶- DnaB 相互作用的频率可反映后随链起始的频率。引发酶也通过 τ 亚基与 DNA PolIII 全酶发生相互作用,这个接触提高了引发的频率并使引物的大小限制在 12 个残基。

四、复制体的组装

引物 RNA 合成后,DNA 聚合酶组装到引发的 DNA 上,完成复制体(replisome)的组装,开始 DNA 合成。聚合酶组装过程包括 γ 复合物催化 β 二聚体结合 DNA,随后 γ 复合物脱离 β 二聚体,PolIII* 与 β 亚基结合完成全酶的装载。

在 PolIII 的 γ 复合物的帮助下,作为滑动钳(sliding clamp)同源二聚体的 β 亚基组装到引发的 DNA 上。滑动钳是一个闭合蛋白质环,需要滑动钳载体把它们组装到 DNA 上,装载过程需要能量。大肠杆菌的滑动钳载体(γ 复合物)含有多个亚基,γ 复合物有较弱的 ATPase 活性,这个活性依赖于 DNA 的存在,并可以被 β 亚基促进。在引发的 DNA 上 PolIII 全酶形成一个引发复合物需要水解两分子 ATP,表明在 DNA 上组装 β 亚基时,每个 γ 水解一个 ATP。在 γ 复合物中与 β 亚基发生相互作用的是 δ。实际上 δ 与 β 之间相互作用不需要 ATP,但由于 δ 隐藏在 γ 复合物中,因此需要 ATP 诱导 γ 复合物构型发

生变化使δ亚基暴露出来与β相互作用，从而体现了对ATP的依赖性。

大肠杆菌的γ复合物催化的基本装载步骤包括：①γ复合物结合ATP导致构型变化，使其中δ亚基上与β亚基结合的位点暴露出来；②δ与β结合并使β开环；③γ复合物识别引物-模板DNA，γ复合物能结合包被SSB蛋白的单链DNA，但不结合裸露的单链DNA，γ复合物识别引发模板可能是通过其χ亚基与SSB蛋白相互作用实现的。γ复合物通过与DNA和β相互作用，使β接近于引物的末端并结合DNA；④ATP水解和ADP释放使γ复合物构型变化，δ亚基不再暴露，从而破坏δ与β之间的相互作用。这导致β亚基闭合形成了一个围绕DNA的钳，以完成滑动钳的装载过程。

PolIII*由14个多肽组成：$\alpha_2\varepsilon_2\theta_2\tau_2\gamma_2\delta_1\delta'_1\ \chi_1\psi_1$。可以视为由PolIII′($\alpha_2\varepsilon_2\theta_2\tau_2$)和γ复合物($\gamma_2\delta_1\delta'_1\ \chi_1\psi_1$)的组合，但PolIII′与γ复合物混合并不能直接生成PolIII*。如图4-17所示，PolIII的各亚基以一定顺序组装，一旦组装完成则很难分开。γ复合物只有在强烈处理情况下才能从全酶上脱落，但是一些γ复合物在细胞中以自由形式存在，这些γ

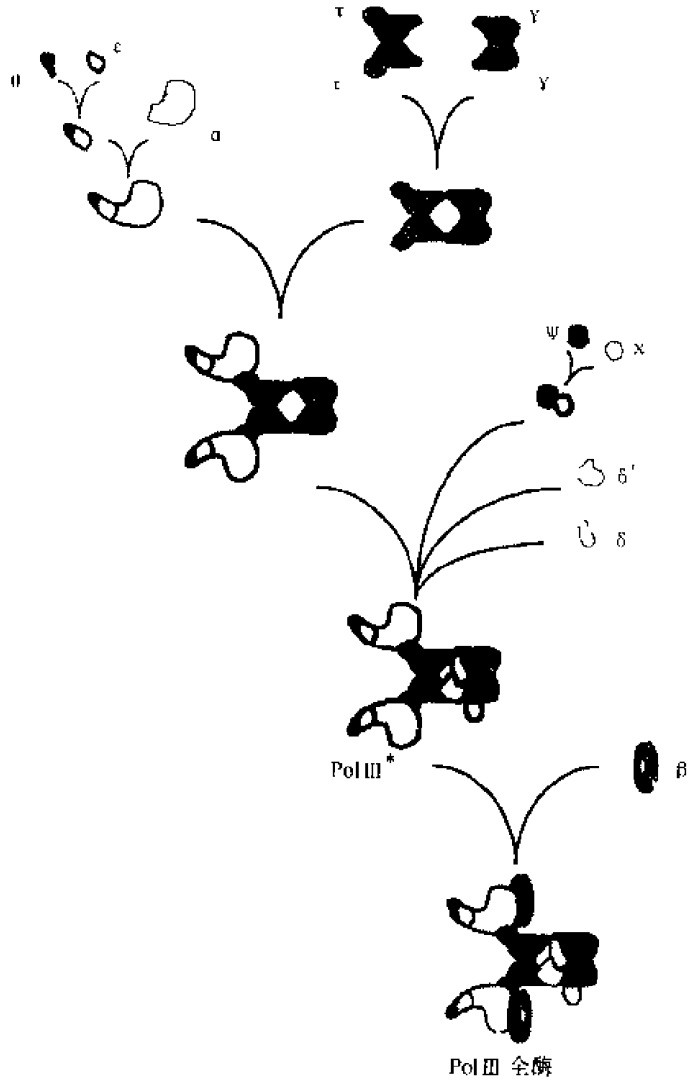


图4-17 PolIII全酶的列组装过程

复合物可能具有其他生物活性。组装 PolIII* 的关键步骤是 τ 与 γ 相互作用形成一个 $\tau_2\gamma_2$ 四聚体，当四聚体形成后，再加入两个核心酶和 δδ′χψ 亚基。如果在 τ 与 γ 混合前 δδ′χψ 加入到 τ 或 γ 上，将抑制 τ 与 γ 的结合，实验表明起抑制作用的是 δ 和 δ′亚基。因为 δδ′χψ 亚基是在 τ 与 γ 结合后加入的，它们在 τ 还是在 γ 上还不清楚，有两种可能：τ 结合核心酶后降低了这些亚基与 τ 的结合能力，而使 δδ′χψ 亚基更易与 γ 结合而形成 γ 复合物；另一个可能是它们位于 $\tau_2\gamma_2$ 四聚体的缝隙上，而且由于 δ′亚基有微弱的依赖 τ 亚基的钳载体活性，因此在 PolIII* 中可能存在两种钳载体，分别由 τδ′和 γδ 构成。

在没有核心酶的情况下，β 二聚体也可以通过 γ 复合物催化结合到 DNA 上。当 γ 复合物使 β 结合到引发的 DNA 上后，很容易脱落到溶液中，脱落的 γ 复合物仍然有活性，它可以通过多次循环使许多个 β 二聚体结合到 DNA 上。利用核心酶、γ 复合物、β 亚基构成的亚聚集体进行研究，发现在延伸过程中 DNA 上只有核心酶和 β 亚基存在，表明 γ 复合物已脱落；而利用全酶进行 DNA 合成时，在 DNA 上 γ 复合物与核心酶、β 亚基共同存在，即 γ 复合物与 β 亚基脱离后仍保持在全酶上。全酶的 τ 亚基作为桥把核心酶和 γ 复合物结合起来，这种排列有利于 β 亚基从 γ 复合物转移到核心酶上，从而完成全酶在复制叉上的组装，开始 DNA 的合成。（图 4－18）。这种连接还能使 γ 复合物的钳载体活性始终保持在复制叉上，对后随链上的多次起始事件是有利的。

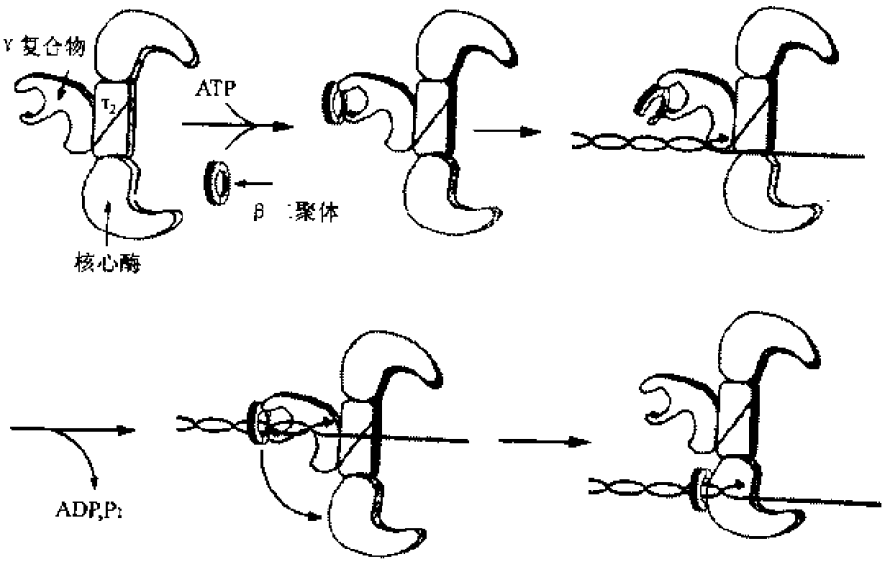


图 4－18　β 亚基从 γ 复合物转移到核心酶上

核心酶和钳载体都能结合 β 滑动钳蛋白表面的相同位点，比较不同细菌编码 β 亚基的基因序列，发现大多数保守残基都位于 β 二聚体的两个 C 末端，而且 C 末端残基发生点突变可使 β 亚基既不能与 γ 复合物相互作用，也不能与核心酶结合，从而失去复制功能，表明核心酶和 γ 复合物都与 β 亚基上这个位点结合。生物体之所以采取这种安排，是因为根据滑动钳装载的机制，γ 复合物既能催化 β 亚基装载又能催化 β 亚基从 DNA 上脱落，这种竞争性排列保证了当核心酶使用 β 亚基延伸 DNA 时，γ 复合物不能卸载 β 亚基

而中断延伸过程，而未与聚合酶结合的自由β亚基可被卸载并重新利用。

在复制体中，全酶的τ亚基还能结合DnaB螺旋酶，这种结合可提高DNA合成速度，在无τ与DnaB结合的情况下，DnaB螺旋酶每秒解旋35 ntDNA，而结合后解旋速度提高了10倍。这种接触也提高了前导链聚合酶的进行性。τ亚基作为一个结构支架在复制体起重要作用，它可促进核心酶的二聚化，是前导链和后随链协同合成的结构基础，并保证螺旋酶、引发酶、聚合酶各部分协同工作。

第四节　延伸过程

DNA聚合酶把新生链的第一个核苷酸加到引物的3′羟基上，从而开始新生链的延伸合成过程。在复制叉前进过程中，必须有DNA螺旋酶解开两条缠绕的模板链。起螺旋酶作用的可能仍是引发体中的DnaB螺旋酶，另一个螺旋酶Rep蛋白也可能参与这个过程。螺旋酶并不能解决复制叉前进所造成的DNA过度缠绕，这必须由拓扑异构酶来解决。在体内正常的环境下，解旋、包被、复制反应是串联进行的。SSB随着复制叉的前进而结合DNA，保持两条亲代链分离，使它们处于合适的环境发挥模板的功能。

在从新起始过程中，两条链的合成并不是同时进行的，一条链（前导链）的合成先于另一条链（后随链）的合成。前导链的合成形成了一个单链DNA的环，当解旋点前进时这个环会变大。因此在第一个冈崎片段合成开始前，存在着一个复制泡，它是由一个双链（一条亲链与前导链配对而成）和一个单链所组成。因为前导链置换了未复制的亲链，所以这个结构称为置换环（displacement loop）或D环，这通常是一个瞬间构型。这种D环与哺乳动物线粒体的D环复制的情况一致，只是这种结构是暂时的，置换的单链上随即进行后随链的合成，而线粒体的D环在复制过程中长期存在，并以单链方式复制。

一、后随链合成的起始

关于后随链冈崎片段的引发机制，人们是利用单链DNA噬菌体的复制作为模型来研究的。因为这些噬菌体在复制的过程中，大部或全部使用宿主的复制酶系和蛋白质，因而研究这些噬菌体DNA的复制对于了解大肠杆菌DNA的复制机制是十分有用的。研究较多的有M13、G4、*ΦX*174三个噬菌体。

在大肠杆菌中有两种类型的引发体系：一种是*ΦX*引发体，以噬菌体*ΦX*174命名；另一种是*oriC*引发体，以细菌复制起点命名。*ΦX*174引发体与*oriC*引发体的组成物质，尽管有一些个别蛋白质不同，但二者都有相同的基本成分，表明噬菌体*ΦX*174的引发与大肠杆菌DNA后随链冈崎片段的起始过程类似。

在*ΦX*174的复制中，引发体在一个特定的引发体组装位点（primosome assembly site，*pas*）组装，这也是*ΦX*174互补链合成的起始位点。引发体包括6种蛋白质：PriA、PriB、PriC、DnaT、DnaB、DnaC。PriA负责识别冈崎片段合成的起始位点（*pas*），并沿着后随链模板的3′-5′方向移动置换SSB蛋白，从而开始引发体其他组分的组装。尽管引发体起始于*ΦX*174 DNA的*pas*，引物却在不同的位点合成，而且引发位点只有一个，而引物合成位点，即冈崎片段的起点有许多个，这表明引发体或它的部分组分可沿着单链DNA移

动到其他位点，引发复制的起始。催化真正的引发反应需要引发酶的参与，*ΦX* 和 *oriC* 复制子的引发酶都是 *dnaG* 基因的产物。DnaG 引发酶短暂地与引发体结合，并受 DnaB 激活起始引物的合成，引物 RNA 的长度被限制为 8－14 个碱基，然后 DNA 聚合酶 III 负责取代引发酶，开始冈崎片段合成。

在 *ΦX*174 复制子中，PriA 水解 ATP 为反应提供能量，它具有螺旋酶活性和置换 SSB 的双重作用。DnaB 是 *ΦX*174 和 *oriC* 中的关键组分，它以 DnaB/DnaC 复合物的形式结合在 DNA 上，DnaB 具有结合引发酶和推动复制叉前进（螺旋酶活性）的双重活性。预引发状态也需要 DnaT 的参与。大肠杆菌 *oriC* 复制不需要 PriA（也包括 PriB 和 PriC）的参与，*oriC* 复制子的引发体只简单地由 DnaB 组成。因此在两种引发体中，DnaB 进入起始位点的方式不同，在 *oriC* 位点和随后的冈崎片段上进行的引发反应中，DnaB 与 DNA 的结合不需要 PriA 蛋白的帮助。我们不能完全理解 *ΦX* 和 *oriC* 起点的差别，但在每个体系中 DnaB 都是冈崎片段前进和引发所需要的关键组分。

二、前导链和后随链的协同合成

半不连续复制的显著特点是：新合成的两条 DNA 链是沿相反方向生长的，两个生长点之间在 DNA 上相隔一段距离，暗示在复制叉附近必然有两个催化 DNA 合成的复合物，分别催化前导链和后随链的协同合成（coordinating synthesis）。前导链的合成方向与复制叉运动的方向一致，因此只需要一次引发事件，PolIII 全酶就能连续不断地延伸前导链。由于存在着把聚合酶束缚在模板上的滑动钳亚基，前导链的合成是高度进行性的。后随链合成方向与复制叉的运动方向相反，因此后随链的延伸是一个不连续的过程，需要重复不断地合成 RNA 引物，再通过 PolIII 全酶延伸成短的冈崎片段。

由于 PolIII 全酶含有两个 DNA 聚合酶亚基，因此在复制叉处前导链和后随链的合成可由同一个 PolIII 全酶分子催化进行，它们协同作用同时合成双螺旋 DNA 的两条链。聚合酶的二聚体结构是双链协同复制机制的基础，两个聚合酶亚基可能相互接近并通过变构效应进行通讯。

在大肠杆菌染色体 DNA 进行复制延伸时，DNA 聚合酶 III 全酶并不是单独起作用，而是与引发体、螺旋酶等构成蛋白质复合体——复制体。由于复制体的存在，前导链和后随链得以同时复制。当复制体沿着 DNA 移动时亲链解旋，前导链连续延伸，引发体在后随链上周期性地起始冈崎片段的合成。

图 4－19 显示了在复制叉处前导链和后随链协同复制的机制。DNA 聚合酶 III 的两个催化核心分别与 DNA 的一条模板链结合。当全酶随着复制叉沿着前导链的合成方向移动时，后随链的模板被甩出来，产生一个环。随着冈崎片段的延长，这一环不断扩大，并向复制叉前进的方向回折。新合成的冈崎片段也向相同的方向（复制叉前进的方向）甩出，这相当于后随链的核心酶相对后随链向与复制叉前进相反的方向移动。因此既保证了两条链的合成方向都是 5′→3′方向，又保证了后随链核心酶的移动方向能够与复制叉移动的方向保持一致。核心酶二聚体在全酶中可能的不对称排列是这个机制的基础。在聚合酶完成一个片段并准备开始下一个片段合成前，这个单链模板必须被延伸至少一个冈崎片段的长度。当冈崎片段合成到前一个冈崎片段的 5′端时，这一大段环就被释放出

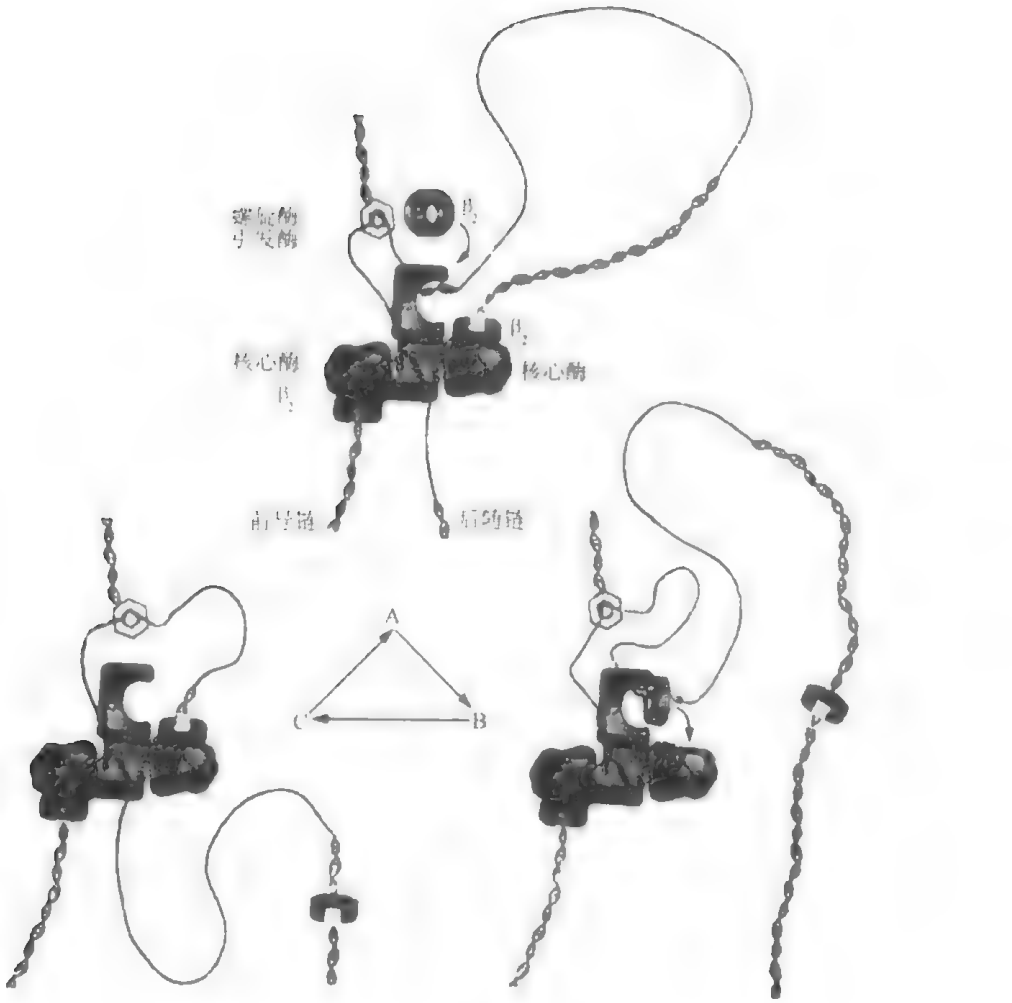


图 4-19　前导链和后随链的协同合成和全酶的循环

来。而前导链的合成又将另一部分后随链的模板置换出来，并在适当的位点由引发体合成新的引物，进行新的冈崎片段的合成。

前导链和后随链的协同合成对于生存是必须的，在 DNA 合成时遇到一个损伤位点将导致一条链合成停止，此时如果另一条链的合成不能相应停止，会造成更大的危害。例如，如果前导链聚合酶连续合成而后随链聚合酶由于某种原因终止合成，后随链的模板就会以 ssDNA 形式不断甩出而暴露在细胞中。细胞中大约有 800 个 SSB 蛋白四聚体，只能覆盖 50 kb ssDNA。当 SSB 蛋白耗尽后，裸露的 ssDNA 将受到核酸酶攻击，而 ssDNA 的断裂很难修复。如果双链协同合成，DNA 损伤因子能导致复制停止，从而避免更大的损失。

三、全酶的循环

由 PolIII 全酶催化的 DNA 合成是一个由多步高效有序的反应所构成的复杂过程。前导链和后随链的合成是不同的，前导链的聚合酶只需要连续地钳住 DNA，而后随链上

是以一系列短的冈崎片段的形式不连续合成的。为了在30～40分钟内完成染色体的合成，聚合酶全酶的合成速度接近1 kb/秒，而冈崎片段的平均长度只有1～2 kb，因此每隔1～2秒就会发生一次冈崎片段链合成的循环。当一个冈崎片段合成完成后，另一个冈崎片段的合成需要从新的位置起始，这个位置位于前导链的生长点附近。如果合成每个冈崎片段都需要一个全酶，每个细胞就需要4000分子全酶进行复制，但全酶是很稀有的，每个细胞中只有10～20个分子全酶，因此当完成一个冈崎片段合成后，聚合酶必须通过一种特殊机制快速循环，转移到新的引发位点上，进行下一个冈崎片段的合成。

为了保证后随链合成能跟得上前导链合成的速度，再循环过程必须十分迅速，只能占整个聚合时间很小一部分，因此需要聚合酶能快速循环。PolIII的再循环与聚合反应高度进行性发生了矛盾。为了实现进行性地合成，全酶通过完全包围双螺旋的β滑动钳亚基紧密钳住DNA，这种紧密结合将阻碍聚合酶的快速循环。全酶利用特殊的机制解决这个问题，每个冈崎片段合成都包括全酶结构的部分去组装和再组装(reassociate)过程。这个过程由ATP水解提供动力，并受专一性的蛋白质－蛋白质和蛋白质－DNA相互作用控制。

当复制体沿着DNA移动时，解旋亲代链，延伸前导链，引发体周期性地起始后随链的冈崎片段合成。当完成冈崎片段合成时DNA复制酶可能有两种去向，它可能从模板上脱落，所以必须有新的复合物重新聚集以延伸下一个冈崎片段；也可能同一个复合物被再利用，如沿着DNA向相反方向转移到下一个引发位点。不同的DNA聚合酶可能采用不同机制来解决这个问题。PolIII全酶的循环过程包括：①DnaB激活引发酶；②引物合成；③在新的引物模板连接处装载β钳；④使DNA PolIII全酶负责后随链合成的核心酶转移到新的引物末端；⑤冈崎片段合成；⑥一个冈崎片段合成后，合成后随链的核心酶从滑动钳(β亚基)上脱落，脱离开已完成的片段，开始下一个冈崎片段的再循环。

聚合酶之所以能够循环到新的引发位点，是因为全酶具有部分去组装，然后重组装的能力，并能在不同的多亚基结构之间快速转化。尽管合成后随链的核心酶在DNA进行性合成时能与β钳紧密结合，但当它到达模板末端并遇到前一个冈崎片段的5′末端时，会突然失去与β钳的亲和力，这时聚合酶会从DNA上快速脱落并使β钳留在后面。这个步骤允许聚合酶循环到下一个引物上，对冈崎片段的合成十分重要。早期人们认为PolIII*从完成的模板上脱落需要DNA分子其他位点上的第二个β滑动钳诱导，但目前研究表明PolIII*从β钳上脱离不需要其他帮助。一旦脱离完成的冈崎片段，PolIII*可循环到上游的新引发位点，快速结合新的β钳。

在复制叉处聚合酶的循环机制与全酶结构相适应，全酶的二聚性质促进了循环过程，因为钳载体和负责后随链合成的核心酶与全酶一起随复制叉移动。在图4－19中，全酶位于运动的复制叉下游(context)，而且每个核心酶都结合一个滑动钳，合成DNA的一条新链，完成双链的进行性延伸。在从A到B的过程中，γ复合物在复制叉新引发位点周围组装β环。同样在A到B中，后随链核心酶完成一个冈崎片段到达DNA上的缺口时，从β滑动钳和DNA上释放，这个释放使核心酶上β结合位点空闲，使核心酶能结合上游RNA引物处新的β钳。在从B到C中，后随链的核心酶循环到新的β钳上，开始下一个冈崎片段的合成。细胞中β亚基的数量比全酶聚集体多20倍，使β可以在前一个冈崎片段完成前，结合到新的引物末端。因此，后随链的聚合酶可在β钳之间转移，而不必从复

制叉处的蛋白质复合物上脱落。

PolIII 复制机器平稳发挥作用需要许多蛋白质-蛋白质和蛋白质- DNA 之间的专一性相互作用。这些相互作用保证冈崎片段合成的每个必须步骤能以适当顺序发生并能够快速完成。快速循环的一个原因是所有的反应组分通过蛋白质-蛋白质相互作用而紧密结合在一起,因此尽管负责后随链合成的聚合酶不断从完成的冈崎片段末端脱落,负责前导链合成的聚合酶在染色体 DNA 复制过程中却始终通过 β 钳紧密地束缚着 DNA,保证了 PolIII 全酶始终结合在复制叉上。聚合酶活性位点与新合成的 RNA 引物在物理位置上的接近也促进了聚合酶的快速循环。同时在全酶中 γ 复合物的存在保证了 β 钳在新合成的引物上快速组装,及它们从完成的 DNA 链上的快速去组装。γ 复合物更靠近负责后随链合成的核心聚合酶,使两个核心酶从 DNA 上脱落的能力不同,分别对应于前导链连续合成和后随链不连续合成的需要。估计整个循环时间远远少于 1 秒,而完成冈崎片段的合成需要 1－2 秒,因而聚合酶再循环所需时间不会影响聚合酶发挥功能。

由于细胞中 β 二聚体的数目与大肠杆菌基因组复制过程中产生的冈崎片段数目相比低一个数量级,所以 β 环也必须重新利用。体外实验表明,PolIII* 的 γ 复合物本身就能催化 β 二聚体从 DNA 上脱落(即钳的载体同时也是钳的卸载体)并用于新的位点。γ 复合物作为载体还是卸载体可能取决于其结合的 DNA 结构,当 γ 复合物结合引物末端时,它将采取一种构型,有效地水解 ATP 并通过协同作用使 β 钳转运到 DNA 上;当结合其他 DNA 结构或在溶液中自由存在时,γ 复合物可能采取不同的构型从而失去了协同作用,这个构型可能催化 β 蛋白环的开关,从而保持 β 钳结合 DNA 或从 DNA 上脱落的平衡。

由于 PolIII 核心酶和 γ 复合物都能通过与 β 环单体的 C 末端相互作用而结合到 β 环的同一个面上,它们结合位点的重叠导致 PolIII 核心酶和 γ 复合物不能同时与 β 钳结合。在 DNA 进行性地合成中,β 钳与 PolIII 核心酶稳定连接,防止 γ 复合物进入,因而有效防止了滑动钳的过早脱落。只有当完成一个冈崎片段合成后,PolIII 的核心酶从 β 钳上脱落时,γ 复合物才可以接近 β 钳并导致钳的脱落。

在细胞浓度的 β 二聚体、DNA、γ 复合物和谷氨酸钾下,每 1/2 秒就可组装一个 β 钳,因此 β 钳的组装足够快,保证了每个冈崎片段上都有一个新 β 钳。这是因为 γ 复合物存在于全酶的结构中,十分接近 DNA,导致复制叉处 γ 复合物的有效浓度非常高。

许多 DNA 聚合酶协同前导链和后随链合成的基本机制都是一致的。最简单的体系是噬菌体 T7,它含有聚合酶亚基、细菌蛋白质硫氧还蛋白(thioredoxin,可增加催化亚基的进行性)以及两个噬菌体基因 4 的产物分别提供引发酶和螺旋酶活性。与 DNA 聚合酶 III 类似,功能性亚基是二聚体,每条 DNA 链上都有一个催化亚基-硫氧还蛋白复合物。实验表明在复制中没有一个亚基脱落成自由状态,排除了合成后随链时聚合酶完全脱落的可能。

T4 也是采用二聚体聚合酶协同完成 DNA 前导链和后随链的同时合成。T4 和真核生物的复制机器都有与 β 二聚体结构类似的环状同聚体组成的滑动钳(分别为 gp45 和 PCNA),而且含有功能上类似于 γ 复合物的滑动钳装载体。

四、冈崎片段的连接

当 DNA 聚合酶遇到前一个冈崎片段的 RNA 引物时,将停止 DNA 的合成。DNA 聚

合酶 I 的 5′-3′外切核酸酶活性切除 RNA 引物，并在 DNA 上产生一个缺口。这个缺口可被 DNA 聚合酶 I 填充，合成一段 DNA 取代 RNA 引物，这种填充反应与切口移位一致，只是新合成的 DNA 取代的是一段 RNA 而不是 DNA。切口移位的长度受 PolI 低进行性的限制。RNaseH 也可以切除大部分 RNA 引物（除了 RNA-DNA 连接处的核苷酸）。当 RNA 被切除并取代后，邻近的冈崎片段必须被连接到一起，一个片段 的 3′羟基末端紧邻前一个冈崎片段的 5′磷酸末端，填平这个缺口的酶是 DNA 连接酶。在原核生物和真核生物中都有连接酶的存在，在 *lig* -突变体中存在未连接的片段，因为没有连接酶把冈崎片段连接到一起。

第五节　复 制 终 止

大肠杆菌染色体复制起始于起点，两个复制叉沿环状染色体以相反方向移动，并在与起点相对的区域终止，但复制不是通过相遇而简单终止的。在复制叉相遇点两边各有两个终止区域，各含有多个位点，分别为 *terE*、*terD*、*terA* 和 *terF*、*terC*、*terB*。这些位点可结合专一性的终止蛋白，引起复制叉制动。多个位点组织成一个复制叉陷井，使复制叉进入终止区域而不能离开。每个终止区域只对一个方向的复制叉有作用。例如在大肠杆菌中顺时针方向的复制叉在遇到 *terC* 停止前，将通过 *terE*、*terD*、*terA* 位点，这些位点对顺时针方向的复制叉没有终止作用。通过这种巧妙的安排，每个复制叉在到达其专一的终点前不得不越过另一个复制叉。如果由于某种原因一个复制叉的运动延迟，导致两个复制叉不能在通常的中心点相遇，稍快的复制叉将被陷于 *ter* 区域，等待较慢的复制叉到来。

所有 *ter* 序列都包含一段短的（22 bp）区域，可在体外终止复制。在许多质粒中也发现了与该终止序列一致的序列，在这些 *ter* 共有的序列附近都可产生终止。但目前还无法确定 *ter* 序列与真正的复制终止位点之间的关系。终止的实现需要 *tus*（terminus utilization substance）基因的产物，这种蛋白质分子量 36 kDa，以单体形式识别 *ter* 共有序列并阻止复制叉行动（图 4-20）。

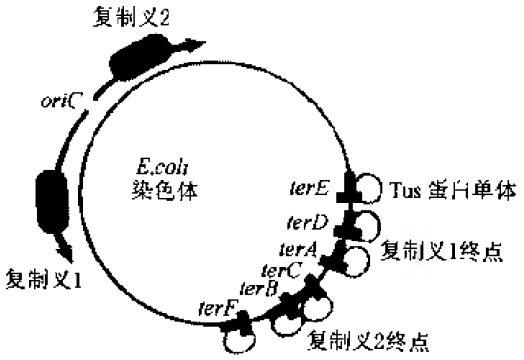


图 4-20　大肠杆菌复制终止区域的结构

双向复制起点为 *oriC*，两个复制叉由 DnaB 螺旋酶代表，终止位点为 *ter*。T 形状代表位点的极性，复制叉遇到平头的一边终止。终止位点的表示不成比例，*terA*、*terD*、*terE* 位于 23～29 分钟之间，*terB*、*terC* 位于 33 和 36 分钟，*terF* 位于 48 分钟。每分钟约为 50 kb，因此这些位点分布了很大的距离。

不同生物体的终止体系是不同的，例如细菌大肠杆菌和枯草杆菌(*B. subtilis*)的终止体系尽管表面类似，但在细节上有很大不同：两个细菌的终止位点没有相关性，而且识别相应位点的蛋白质也不同。枯草杆菌的终止位点是 47 bp 的不完全反向重复序列，由复制终止子蛋白(replication terminator protein，RTP)识别。RTP 单体分子量为 14.5 kDa，可形成稳定的二聚体，而且两个二聚体同时作用于每个终止位点。Tus 和 RTP 几乎没有序列同源性。

一、大肠杆菌 *ter*-Tus 复合物

Tus-*terB* 之间可形成专一性的复合物($k_0 = 3.4 \times 10^{-13}$)，这个复合物非常稳定，其半衰期为 550 分钟。DNA 保护实验表明蛋白质与 DNA 以一种不对称的方式结合，在复合物靠近复制叉终止位点的一侧，DNA 的双链都与蛋白质接触；而在另一侧，只有一条链上的碱基与蛋白质作用。

在体外，*ter*-Tus 复合物(*ter*-Tus complex)可在大肠杆菌染色体复制起点阻止复制叉运动；在体内，只有当复制叉以正确的方向进入终止位点时，才能发生复制终止。除了 Tus 蛋白(terminus utiligation substance)外，终止不需要其他蛋白质，复制停止于 *ter* 序列的第一个核苷酸。

ter-Tus 复合物可能通过抑制 DNA 螺旋酶来发挥终止作用，螺旋酶利用水解 ATP 的能量融解 DNA 双螺旋。在螺旋酶单独催化的解旋反应中，如果螺旋酶的 DNA 底物存在 *ter* 位点，将会抑制解旋反应；抑制作用也有方向性，与完整复制叉的情况相同。DnaB、PriA 及含有 DnaB 和 PriA 的 ΦX 引发体都能被抑制。UvrD 蛋白(螺旋酶 II，参与 DNA 修复)和 Rep(参与一些噬菌体复制的大肠杆菌螺旋酶)在一些研究中有抑制效果，而在另一些研究中则没有。大肠杆菌螺旋酶 I、SV40 T 抗原、螺旋酶 B 也可被 *ter*-Tus 复合物抑制。

这些研究表明在 *ter*-Tus 复合物中蛋白质-DNA 相互作用抑制了螺旋酶。另一些实验表明在 Tus 介导的终止中存在更微妙的作用。例如 *ter*-Tus 复合物抑制螺旋酶的能力受实验环境的影响。当需要解旋的 DNA 片段变长时，与短的底物相比，DnaB 和 PriA 对 *ter*-Tus 的抑制不敏感。但含有多种螺旋酶活性的 ΦX 引发体却可被 *ter*-Tus 复合物抑制。显然这些螺旋酶在多蛋白质复制复合物中比单独存在的螺旋酶对终止更敏感。*ter*-Tus 复合物也能阻碍由大肠杆菌 DNA 聚合酶 I 和噬菌体 T5、T7 DNA 聚合酶催化的链置换合成。在两个方向都可发生抑制，而正确方向的位点的抑制作用更大。但与完整的复制叉相比，DNA 聚合酶可更深入 *ter* 位点。因此，在抑制聚合酶和抑制螺旋酶两种可能性中，*ter*-Tus 复合物对螺旋酶的抑制更类似复制终止的情况，表明螺旋酶是大肠杆菌终止复合物的靶子。

二、抑制机制

不同细菌的终止体系有相同的基本性质：第一，专一性的蛋白质-DNA 相互作用足以阻碍一个方向的复制叉；第二，复制叉阻碍在 DNA 终止位点的边缘，这个位点被结合

蛋白所保护;第三,终止蛋白以不对称的方式结合终止位点,在复合物中,阻碍的一侧比非阻碍的一侧有更多的蛋白质-DNA 相互作用;第四,复合物抑制大肠杆菌 DnaB 螺旋酶,这是大肠杆菌染色体复制时负责复制叉运动的主要螺旋酶。基于 *ter*-Tus 复合物阻碍许多蛋白质在 DNA 上运动的能力,可以认为它在 DNA 上形成了一个单向的关卡,阻止蛋白质通过。当终止蛋白发挥关卡作用时,复制叉和终止复合物之间专一性的蛋白质-蛋白质相互作用也可能产生终止功能,目前不能确定 DNA-蛋白质和蛋白质-蛋白质相互作用对终止机制的贡献程度。

根据细菌染色体终止位点的排列,如果两个复制叉到达其相应的终止位点前都不停止运动,两个终止信号之间的区域将被复制两次。但许多体内和体外实验表明,每次复制时只使用了一个终止位点。显然,当一个复制叉首先遇到其相应的终止位点时,将停止运动,等待另一个复制叉到来。相反方向的复制叉也在同一个位点停止,并不继续前进到其自身的终止位点。首先在相应终止复合物处停止的复制叉会形成相反方向复制叉的障碍,保证每个染色体片段精确复制一次。

在终止复合物处复制叉明显暂停,然后释放其蛋白质组分。目前尚不清楚终止复合物除了停止复制的作用,是否还有去组装复制叉的作用。

三、复制终止体系的生物学功能

研究表明终止体系功能是防止 DNA 过度复制。当不存在终止体系时,一轮复制完成后,DnaB 螺旋酶不能停止运动,它将继续解旋由模板链和新合成链组成的双螺旋 DNA。同时聚合酶继续延伸第一轮复制产生的新链的 3′末端,从而产生长的 DNA 多聚体。*ter*-Tus 复合物终止 DnaB 螺旋酶,以防止过度复制,避免出现不正常的多聚体 DNA。对低拷贝数的质粒来说,防止复制时产生 DNA 多聚体至关重要。因为质粒二聚体可能被拷贝数控制机制视为两种质粒,在分裂时不能正常分配到两个子代细胞中。

终止机制也可能是一个起始后调控复制的方法。通常复制调控主要发生在起始阶段,但在一些特殊情况下,如存在过量起始因子和氨基酸饥饿时,可发生不受控制的起始。在这种情况下,终止蛋白能在复制叉到达正常终点前阻止复制叉运动,导致复制中断,避免过度复制。但复制体系如何防止终止蛋白影响正常复制呢?人们发现大肠杆菌编码一个蛋白质,能拮抗 Tus 的终止活性,这个发现从另一方面也支持了终止体系对 DNA 复制的调控作用。尽管目前人们还不清楚这种蛋白质在生物体内的功能,但复制叉在环境刺激下停止并释放的假设是很有可能的。

此外,终止序列对 DNA 重组也有一定影响。一些同源重组在终止区域发生的频率比其他区域高 10~100 倍,可能复制叉停止在终止位点,有助于帮助重组所需要蛋白质的组装。

四、线性 DNA 的末端复制

由于 DNA 聚合酶不能从头起始 DNA 的合成,其复制反应需要 RNA 引物的参与,当 RNA 引物水解后,需要 DNA 聚合酶来填充缺口。但已知的核酸聚合酶都是 5′→3′方向

进行合成的。在线性 DNA 的末端复制时将产生问题，两个子代分子中新合成链的 5′端不能合成。如果这个问题得不到解决，在连续的复制循环中 DNA 链将会变短，造成遗传信息的丢失(图 4－21)。所以必然存在一个特殊的机制来合成线性 DNA 的末端，可能有 4 种解决办法：

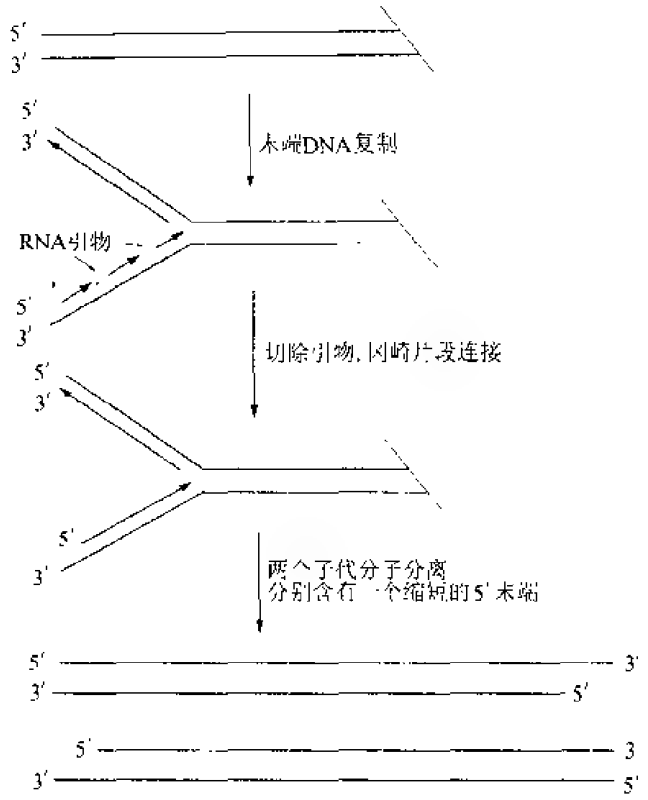


图 4－21　线性 DNA 的末端复制问题

1) 有的线性 DNA 分子从一开始就采取环化的方法。例如 λDNA 在进入宿主细胞以后就环化并成为双螺旋形式。

2) 有的线性分子采取形成连环分子(concatemer)方法。许多线性分子的两端都有一段重复的核苷酸序列，称末端冗余(terminal redundancy)。它意味着两个子代分子的单链末端可互补，然后由 DNA 连接酶连接起来。这样形成的二聚体称连环分子。生成的二聚体还可通过复制产生四聚体、八聚体连环分子。噬菌体 T4 采用这种机制(图 4－22)。

3) DNA 通过在末端产生一个发夹，可形成一特殊结构，导致事实上没有自由末端。草履虫(*Paramecium*)的线性线粒体 DNA 的复制就产生了这种交联的结构。

4) 最直接的解决办法是通过引入一个蛋白质直接从末端起始 DNA 的合成，一些线性病毒的核酸的 5′末端可共价连接蛋白质，借助蛋白质的介入来解决线性分子末端复制问题。腺病毒(adenovirus)DNA、噬菌体 Φ29 DNA、脊髓灰质炎病毒(poliovirus，一种单链 RNA 病毒)RNA 就采用这种方式。

线性 DNA 末端起始复制的主要例子来自腺病毒 DNA 和 Φ29 DNA，它们可以利用

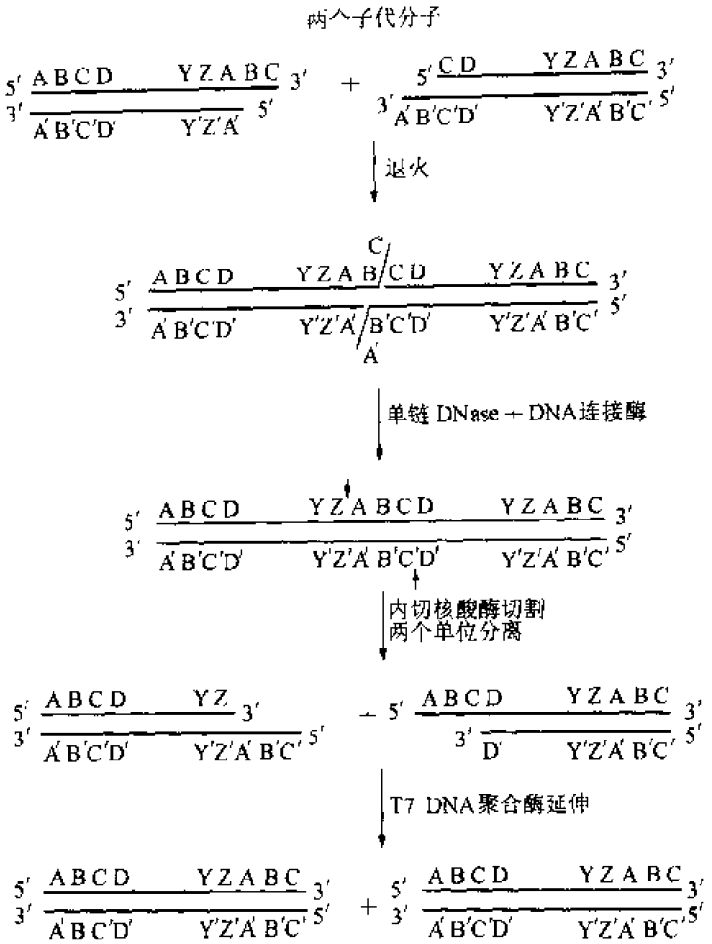


图 4-22 连环分子形式的终止

如图 4-23 所示的链置换方式真正地复制两个末端。新链的合成从一个末端开始，置换了双螺旋 DNA 上的同源链。当复制叉到达分子的另一端时，被置换的链以一条自由的单链形式释放出来，然后独立地复制。这需要在分子的末端以一些短的互补序列通过碱基配对形成双螺旋片段。这些病毒中，都存在可以共价地结合到 5′末端上的蛋白质末端蛋白(terminal protein，TP)。在腺病毒中，是一个分子量 55 kDa 的蛋白质，通过其丝氨酸残基形成磷酸二酯键与成熟病毒 DNA 连接。相同的情况可在 Φ29 中发现。在脊髓灰质炎病毒中，一个只有 22 个氨基酸的 VPg 蛋白通过酪氨酸上的羟基连接到 5′末端碱基上。这些蛋白质都由病毒编码。

与正在复制的腺病毒新生 DNA 链的 5′端相连接的是分子量 80 kDa 的蛋白质，是成熟 DNA 上结合的 55 kDa 蛋白质的前体，称为末端蛋白质前体(pTP)。80 kDa 的 pTP 参与了腺病毒 DNA 的复制起始，在病毒成熟的某个阶段，水解为 55 kDa 的蛋白质，并保留在成熟的病毒分子中。

pTP 蛋白质可与一个 140 kDa 的腺病毒 DNA 聚合酶形成一个复合物。聚合酶可催化 pTP 的丝氨酸残基与 dCTP 之间的反应，使 dC 与 pTP 的丝氨酸残基共价连接。腺病

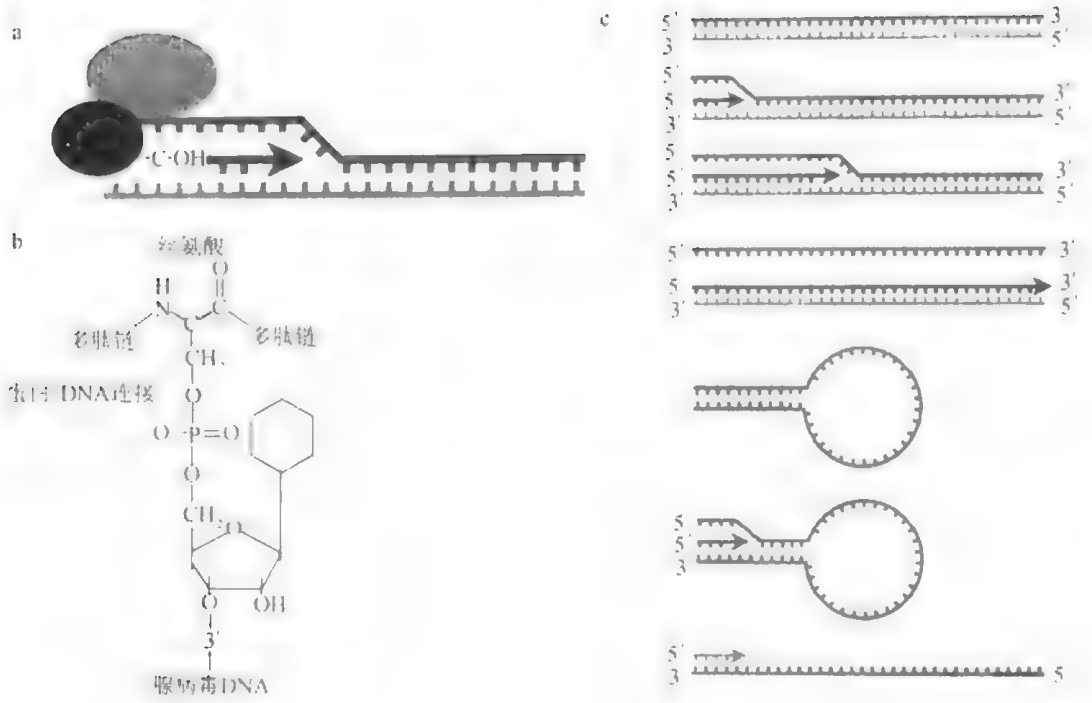


图 4-23　腺病毒 DNA 末端的复制模型

毒 DNA 分子的 5′端连接着 55 kDa 的 TP 蛋白，TP 蛋白可被 80 kDa 的 pTP 蛋白取代，使聚合酶和末端蛋白质的复合物附着在腺病毒 DNA 的末端上。同时 C 与模板链 3′端的 G 以氢键结合，在宿主蛋白 NFI 的作用下，在 DNA 末端的特定区域形成了起始复合物。然后在 C 的自由 3′羟基末端上由 DNA 聚合酶引发腺病毒 DNA 的复制，产生了一个与 pTP 蛋白相连接的新链。相似的机制也适用于 Φ29 复制。

整个反应是协同进行的，pTP 蛋白与 dCTP 的连接是由 DNA 聚合酶在腺病毒 DNA 的存在下完成的，表明连接反应类似于核苷酸-核苷酸之间的连接反应。复制可起始于任何一个末端，但每次只有一个末端形成起始复合物。通过链置换的方式合成 DNA，产生了一个双链分子及一个单链分子。被置换出来的单链的两个末端通过互补序列的碱基配对生成含有一段双链的茎环结构，然后在双链末端以同样的方式合成其互补链。

第六节　其他原核生物的复制体系

一、噬菌体 λ 的复制

噬菌体 λ 的复制需要两个病毒蛋白质的参与，它们是噬菌体 λ 基因 *O* 和 *P* 的产物，还包括除 DnaA 和 DnaC 外的所有大肠杆菌复制蛋白。研究表明，O 蛋白和 P 蛋白分别是 DnaA 和 DnaC 的类似物。复制起点位于基因 *O* 内部，包括三个部分：4 个高度保守的正

向重复的 18 bp 序列——*ori* 重复序列，其中每个 18 bp 序列都是反向重复；约 40 bp 的富含 A/T 的序列；28 bp 的回文序列。携带 *ori*λ 的质粒体外复制时只需要其中两个重复以及富含 A/T 的序列。

噬菌体 λ 中，起始蛋白是噬菌体编码的 O 蛋白，分子量为 33.7 kDa。O 蛋白以二聚体形式结合在 4 个 *ori* 重复序列上，8 个 O 蛋白在 *ori*λ 上形成了一个多蛋白质的复合物。在超螺旋质粒中，O 蛋白与 *ori*λ 的结合导致了 DNA 结构的明显变化，包括 DNA 上富含 A、T 的区域的解旋，对 O 蛋白来说解旋过程不需要 ATP 的参与，但是在大肠杆菌的 *ori*C 上，大肠杆菌 的 DnaA 起始蛋白结合并引起 DNA 的结构变化则需要 ATP 的参与。在两种情况下，组装的核酸蛋白质复合物都作为一个装载体，实现一系列复杂的蛋白质-蛋白质，蛋白质- DNA 之间的相互作用。

DNA 复制起始的下一个重要的步骤是在 DNA 上装载 DNA 螺旋酶，以促进亲代双螺旋的解旋。在噬菌体 λ 中感染的细胞中，噬菌体编码的 P 蛋白与结合 DNA 的 O 蛋白相互作用，促进 DnaB 螺旋酶结合在 *ori*λ 上。随着 O 蛋白聚集体的形成，P 蛋白和 DnaB 螺旋酶以 P-DnaB 复合物的形式加入进来，形成 *ori*λ-O-P-DnaB 复合物。因为 P 蛋白能抑制 DnaB 螺旋酶的 ATP 酶活性，因此 *ori*λ-O-P-DnaB 复合物在随后的一系列复制事件中保持无活性的状态，只有 P 蛋白失活或者从复合物中脱落的情况下，DNA 复制才能起始。

三种大肠杆菌的热激蛋白，DnaK、DnaJ 和 GrpE 的结合促进了 *ori*λ 的活化。它们协同作用从 *ori*λ-O-P-DnaB 复合物上去除 P 蛋白，使 DnaB 活化，从而使 DNA 解旋，这个过程同时伴随着 SSB 的结合。随着 DnaB 螺旋酶的活化，下一个重要步骤是使 DNA 引发酶组装到 DNA 上以产生 DNA 聚合酶合成所需要的 RNA 引物。

二、噬菌体 T7 复制起始

早期研究表明噬菌体 T7 复制起始时在 17% 位置形成复制泡，并双向复制。现在发现其复制起始于 14.75%到 15%的 200 bp 区域。*ori*T7 位于基因 *1* 和基因 *1.1* 之间，包含基因 *1.1* 的两个启动子（*Φ1.1A* 和 *Φ1.1B*）和一个 61 bp 富含 A/T(79%)的区域，富含 A/T 区域的模板链上有 T7 基因 *4* 蛋白的引发位点。复制起始需要 T7 RNA 聚合酶、T7 DNA 聚合酶、基因 *4* 蛋白的参与。

T7 RNA 聚合酶是基因 1 的产物，是分子量 98.1 kDa 的单体蛋白质。其启动子含有 23 bp 的保守序列。T7 DNA 聚合酶由两个亚基组成。一个是 T7 基因 *5* 蛋白(g5p)，分子量 80 kDa，具有低进行性的聚合酶活性(1～50nt)，也具有 3′→5′外切核酸酶活性。另一个亚基是宿主 *trxA* 基因编码的硫氧还蛋白(thioredoxin)，分子量 12 kDa，是 T7 DNA 聚合酶的进行性因子。它们形成 1:1 的复合物，具有较高的进行性，可连续合成数千 nt。基因 *4* 蛋白(g4p)有两种分子形式，一种分子量为 56 kDa，另一种分子量为 63 kDa。56 kDa 的蛋白质是从基因内部翻译的产物，与 63 kDa 的 g4p 相比，它缺少 7 kDa 的 N 末端区域。基因 *4* 蛋白是多功能的酶，两种形式的 g4p 有相同的 TTPase 和螺旋酶活性，而 63 kDa 形式的 g4p 的引发酶活性比 56 kDa 的 g4p 高 150 倍。两种形式的酶协同作用时，活性还能提高 100 倍。表明两种形式的 g4p 形成异二聚体时，才能发挥更好的引发作用。当合成体系受到稀释时，前导链的合成不受影响，而后随链的引发频率下降，表明基

因 *4* 蛋白的引发酶活性不是进行性的，而是分散的。

T7 RNA 聚合酶可利用 *Φ1.1A* 和 *Φ1.1B* 启动子中的任何一个发动转录，当转录产物延伸到富含 A/T 区域的某个位点时，将被 T7 DNA 聚合酶作为引物合成 DNA，转录产物长度为 10～60 nt。这条 DNA 链即为前导链。由于转录作用，导致启动子下游富含 A/T 区域的 DNA 融解，使 T7 基因 *4* 蛋白得以进入并结合到非模板链上。T7 基因 *4* 蛋白沿 5′→3′方向移动，解旋双螺旋直到第一个引发位点。在这个位点上，T7DNA 聚合酶延伸引物 3′末端，建立了右向的复制叉。右向复制叉的第一个冈崎片段向左延伸，将转化为向左移动的前导链。随后 SSB 向左运动使 DNA 变性，暴露另一模板链的 T7 基因 *4* 蛋白引发位点，T7 基因 *4* 蛋白进入，提供螺旋酶活性，并引发冈崎片段合成，形成左向的复制叉。

三、噬菌体 T4 的复制体系

T4 DNA 的合成由一系列蛋白质组成的多酶聚集体催化完成。基因 *32* 蛋白是一个高度协同性的单链结合蛋白。T4 复制专一性地需要噬菌体编码的单链结合蛋白，不能被大肠杆菌 SSB 取代。gp32 与 T4 DNA 聚合酶形成复合物，它们之间的相互作用对复制叉的构建十分重要。

基因 *41* 蛋白和基因 *61* 蛋白协同作用，催化形成短的引物。gp41 类似大肠杆菌的 DnaB 蛋白，是一个螺旋酶，具有依赖单链 DNA 的 GTPase 活性，以多聚体的形式发挥功能。基因 *59* 蛋白可促进 gp41 与 DNA 的结合。基因 *61* 蛋白有引发酶活性，类似大肠杆菌的 DnaG 蛋白。引发酶识别模板序列 3′-TTG-5′，合成具有 pppApCpNpNpNp 序列的五核苷酸引物，并在此基础上延伸 DNA 链。gp61 的含量很少，每个细胞只有 10 个拷贝。

基因 *43* 蛋白是 DNA 聚合酶，分子量 110 kDa，它具有 5′→3′合成活性，和 3′→5′校正外切核酸酶活性，能催化 DNA 合成并去除引物。gp43 单独存在合成活性很低，高活性的合成需要其他三个蛋白质的参与。这三个蛋白质作为聚合酶的辅助因子，可提高聚合酶对 DNA 的亲和性，提高聚合酶的合成速度和进行性；基因 *45* 蛋白是二聚体，它的功能类似于大肠杆菌 Pol III 的 β 滑动钳。基因 *44* 蛋白和基因 *62* 蛋白形成一个紧密的复合物，具有 ATP 酶活性，它们的功能可能类似于 γδ 复合物。

此外，基因 *39*，*52* 和 *60* 可编码 T4 拓扑异构酶 II 的三个亚基，可去除复制引起的超螺旋。

T4 复制起始也需要宿主的 RNA 聚合酶，从二个启动子 $P_{ori}1$、$P_{ori}3$、$P_{ori}4$ 的任何一个起始转录，利用合成的 RNA 作为引物，在 *oriA* 位点起始左向的前导链合成，并取代非模板链。在新的前导链 70 nt 内对应的区域含有三个 T4 gp61 的引发位点。T4 引发体可在此进入 DNA，形成向左运动的复制叉。目前还不清楚右向复制叉是如何形成的。另一个启动子 $P_{ori}2$，位于另一条链上与 $P_{ori}4$ 重叠的位置，可能以相同的机制负责右向复制叉形成。

第七节　其他形式的复制

一、线粒体的 D 环复制

大多数生物染色体 DNA 的复制是双向的而且是对称的，然而也有例外，线粒体的 DNA 采取一种特殊的单向复制方式称为置换环或 D-环(D-loop)复制。

脊椎动物线粒体大小 16～18 kb，并以闭合环状分子的形式存在。mtDNA 的两条链在变性的氯化铯离心时迁移率不同，这因为一条链大部分是 G+T 组成，因此命名为重链(H 链)，相应地另一条链为轻链(L 链)。每条链可从一或两个启动子起始转录，根据转录模板所在链的不同，这些启动子被定义为轻链启动子和重链启动子。每条链都产生大的多顺反子的转录产物，然后加工产生成熟的 tRNA、rRNA 和 mRNA，以实现线粒体遗传信息的表达。在目前已检测的脊椎动物中，mtDNA 分子都包括一个非编码区域，其中含有起始转录的启动子和一个紧密相连的重链 DNA 复制起点(O_H)。两条链的合成是不对称的，两条亲本链中只有一条(L 链)可作为新链合成的模板，合成一段重链置换了原初的亲链(H)，使之保持单链状态。如图 4-24，这段发生置换的区域称为置换环或 D 环，这个环是由于模板指导的重链 DNA 合成在起始不久就终止而形成的。这个停止合成的新重链仍然稳定地与环状亲代分子杂交，通过置换亲代的 H 链产生一个三链结构。

在哺乳动物的线粒体中 D 环常常保持 500～600 碱基的长度，新合成的短链是不稳定的，它经常被降解又再次合成，以保持一定长度的开放区域。一些线粒体 DNA 如爪蟾(*X. laevis*)DNA 具有一个很长的 D 环，另一些线粒体则具有多个 D 环，在四膜虫的线性线粒体 DNA 上有多达 6 个 D 环。在叶绿体 DNA 上也有这种现象，有些植物的叶绿体上有两个 D 环。

1. H 链复制的起始

线粒体 DNA 的复制与核 DNA 采取相同的起始方式，首先合成一小段 RNA 引物，然后 DNA 聚合酶在其上延伸。在脊椎动物线粒体中 DNA 复制和转录是相关的，起始子轻链启动子(light-strand promoter，*LSP*)的转录产物是重链复制起点处 DNA 复制的引物。

哺乳动物 mtDNA 通过不对称的置换机制复制，这需要两个方向复制起点的参与。首先在 O_H 处起始 H 链的复制，这个位点在基因组 D 环区域中，位于 *LSP* 下游。哺乳动物线粒体启动子由两部分组成，一部分是转录起始区域，另一部分是转录起始因子 mtTFA 的结合部位。mtTFA 分子可结合 DNA 并使 DNA 弯曲并解链。在所有检测的脊椎动物中，在体外从线粒体启动子起始的转录需要线粒体 RNA 聚合酶组分和 mtTFA。真核生物如酵母甚至人的 mtRNA 聚合酶基因与细菌噬菌体 T3、T7 和 SP6 的 RNA 聚合酶同源。

2. 在 O_H 形成稳定的 RNA-DNA 杂交结构

不论线粒体启动子上转录是如何起始的，*LSP* 位于 O_H 直接上游位置的事实暗示这个启动子在 mtDNA 复制时的作用。通过精确确定 D 环上 RNA 和 DNA 的位置表明从

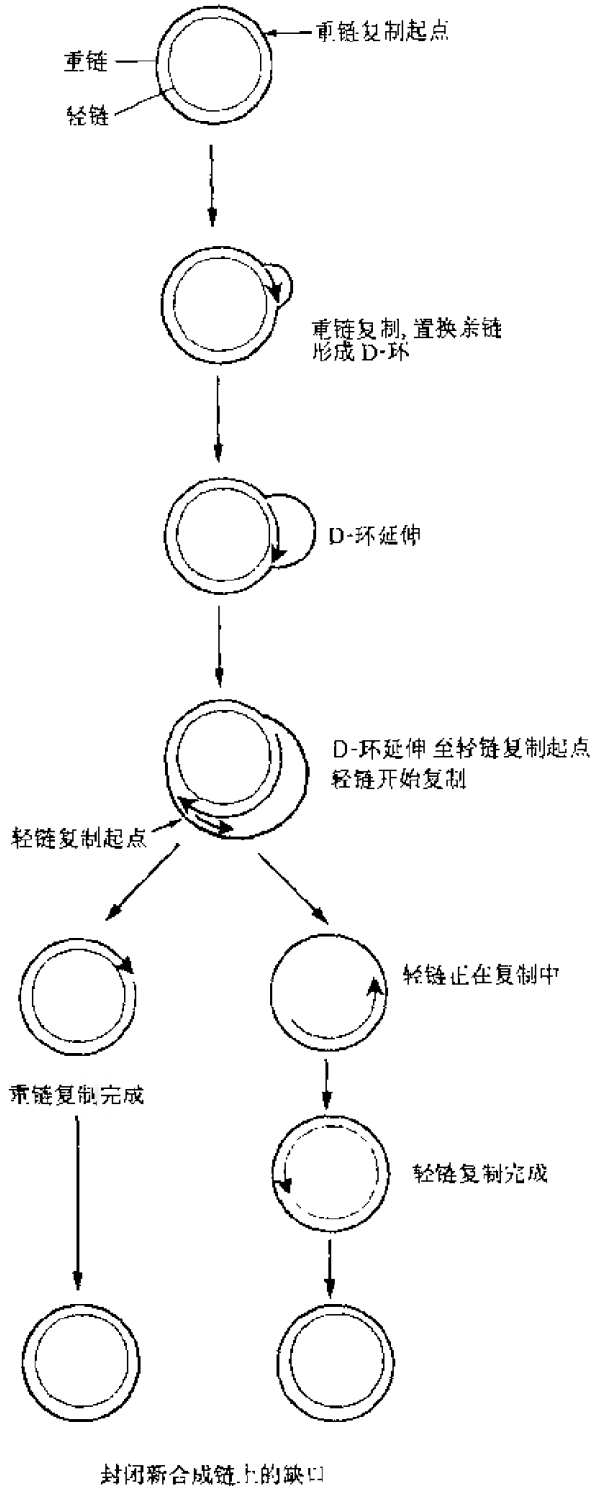


图 4-24 线粒体的 D 环复制

LSP 产生的 RNA 可作为 H 链 DNA 复制的引物。比较人、小鼠、大鼠 *LSP* 下游的核酸序列发现了三个保守的序列框(CSBI、II、III)在所有检测的脊椎动物的 D 环区域,CSB 是保

守的，有证据表明 CSB 可能参与形成一个适当形状的 RNA 底物来加工线粒体引物 RNA。

在酿酒酵母的 mtDNA 中，*ori/rep* 序列包括一个类似 CSBII 的元件，位于转录启动子的下游。在体外实验中用含有 *ori* 启动子的模板转录可产生保持与模板结合的转录产物。这个产物对 RNA 酶 A 和 RNA 酶 T1 有抗性，而对 RNA 酶 H 敏感。形成这种 RNA-DNA 杂交需 CSBII 序列，并且杂交不会发生在含其他线粒体启动子（如 tRNA、mRNA、rRNA 启动子）的模板上。杂交发生在酵母 *ori/rep* 序列上表明 RNA-DNA 杂交对于酵母 mtDNA 复制是十分重要的。与在酵母中类似，依赖于 CSBII 的 RNA-DNA 杂交也形成于人 *LSP* 下游的 CSBI、II 位点，CSBIII 序列的突变也使杂交消失。利用鼠 O_H 进行的哺乳动物 RNA-DNA 杂交也表明存在这样一个特殊的结构，所以 RNA-DNA 杂交是酵母的 *ori/rep* 和人，鼠 mtDNA 的 O_H 序列的保守性质。这个结构可能含有极端稳定 R 环（含有两个 DNA 链和一个 RNA 链的分子，其中一个 DNA 链被杂交的 RNA 分子所取代）。然而，利用酶切和化学修饰等方法检测表明 RNA-DNA 杂交体是一个包括所有 3 条链共同参与的奇特结构，而不仅仅是简单的 R 环结构。在 ColEI 和 T7 原核复制起点的 DNA 复制时也形成短的稳定的 RNA-DNA 杂交，但是线粒体的情况在真核生物中还是第一次发现。

3. 引物 RNA 的加工

为了形成合适的复制引物，*LSP* 转录产物必须经过加工使 3′末端在 O_H 处终止。研究人员在鼠和人的细胞中发现了一种位点专一性的线粒体 RNA 外切核酸酶（RNA 酶 MRP）可在特殊位点加工含有 O_H 序列的 *LSP* RNA 转录产物。此酶含有蛋白质亚基和 RNA 亚基。现已在许多有机体中发现 RNA 酶 MRP 的活性并鉴定了编码 RNA 组分的基因，酶蛋白和 RNA 组分的结构在进化中是保守的。但用含有起点序列的单链 RNA 作为底物时，并不能断裂已知的 RNA-DNA 转化位点。这个矛盾可能是由于 RNA 酶 MRP 加工的底物是三链 RNA-DNA 杂交物，而不是单链 RNA。这些断裂完全依赖于 CSBI 的存在。RNA 酶 MRP 在脊椎动物线粒体的 RNA 引物加工过程是必须的，也进一步证明 RNA-DNA 杂交体是 RNA 加工的底物。

确定线粒体中存在此酶是很困难的，因为在核中也发现了相似的酶活性。有证据表明核 RNA 酶 MRP 参与了 5.8 S rRNA 的晚期加工。通过原位杂交发现少量 RNA 酶 MRP RNA 位于鼠和爪蟾卵母细胞的线粒体中。因此细胞中似乎含有两种 RNA 酶 MRP，在核中大量存在用于 rRNA 加工。在线粒体中少量存在，参与线粒体 RNA 引发。

4. H 链 DNA 合成的终止

哺乳动物线粒体 DNA 复制起始后，新合成的短链继续延伸，导致亲本 H 链上被置换的区域变长，扩展了 D 环。在脊椎动物 mtDNA 中广泛存在三螺旋 D 环，这表明它对于 mtDNA 的代谢具有关键的作用。可能包括保持一定数目的引发 mtDNA 分子以引发随后的 DNA 复制，并提供转录调控和 mtDNA 分配必需的 mtDNA 构型。D 环的形成是由于 O_H 下游 H 链合成的过早终止。这个终止的新 H 链还保持与亲链双螺旋的结合。不同种类 O_H 下游 D 环的 3′末端的位置不同，即 D 环的物理大小具有种属专一性。尽管如

此，D环的3′末端终止于特定的位点，这个位点是由短的模板序列决定的。这些末端结合序列在脊椎动物中是保守的，而且在已检测的D环中位于3′末端上游相似的位置，表明新合成的D环链的终止是一个模板指导的事件，可能对线粒体D环的形成有潜在的调控作用。通过对小牛的末端结合序列结合活性的鉴定，表明新合成的H链是终止还是继续进行H链复制是由核编码的因子决定的。

5. L链复制的起始

在脊椎动物的mtDNA中，轻链复制的起点 O_L 距离 O_H 位点很远，位于mtDNA分子上离 O_H 大约2/3基因组距离的位点。它是由tRNA基因隔开的一段非编码区域。当复制反应沿着线粒体环延伸到 O_L 时，被置换的H链上的起始位点暴露出来，在这个起始位点上L链的合成起始。利用被置换的亲代H链作为模板，以与子代H链合成相反的方向合成L链。亲代H链暴露并以单链形式存在时，这个区域的DNA序列可能形成一个稳定的茎环结构。L链合成的起始需要引发酶产生一段短的RNA分子，其5′末端位于茎环结构富含T的部分。在体外有效的复制不仅需要茎环结构，还依赖茎上的碱基序列，以固定DNA聚合酶延伸需要的引物。在大多数脊椎动物编码Trp、Ala、Asn、Cys、Tyr等tRNA基因的mtDNA序列中，发现存在具有保守二级结构的非编码区域，表明这些序列对于 O_L 的功能是必需的。然而，系统发育比较显示鸡、海鳗、鳄鱼、鳄蜥中未发现类似 O_L 的非编码区域，可能 O_L 位于染色体的其他位置，或其中的一个tRNA分子具有 O_L 的功能。

由于起步较晚，当H链的合成停止时，L链的合成才完成1/3。这时产生了两个环状DNA分子：一个是完整的双链环状DNA，另一个是含有部分单链有缺口的环，将继续L链的合成。合成完成后需要除去两个起点的RNA引物，填充并连接分子上的缺口，并在共价闭合环状mtDNA中引入超螺旋。DNA复制的最后阶段包括两条子代分子的分离。

6. DNA聚合酶γ和辅助蛋白

在脊椎动物中，只有一种催化线粒体DNA合成的DNA聚合酶：DNA聚合酶γ(Polγ)。它具有与其他的细胞DNA聚合酶不同的性质，例如在体外可利用RNA作为模板；被顺丁烯二酰乙胺和ddNTP抑制，对蚜栖菌素(aphidicolin)有抗性，可被盐活化。各种来源的聚合酶γ纯化时都伴随着3′→5′外切核酸酶活性，对应于聚合酶γ反应的高度忠实性。高度纯化的聚合酶γ制品通常包括两种主要的多肽，一个大催化亚基具有DNA聚合酶活性；一个小亚基的功能未知。不同种类有机体编码聚合酶γ的基因已分离纯化。第一个聚合酶γ亚基的基因是从酵母中分离的，编码一个140 kDa的多肽。氨基酸顺序表明这个蛋白质含有与原核生物DNA聚合酶的核苷酸聚合酶和外切酶区域同源的区域。外切酶区域突变在体内导致mtDNA的突变率升高，表明这个区域对错配核苷酸有校对功能。其他证据也表明脊椎动物聚合酶γ的大催化亚基同时具有5′→3′聚合酶和3′→5′外切酶活性。

与核DNA聚合酶一样，聚合酶γ也需要辅助因子的参与。一些有机体的线粒体含有单链DNA结合蛋白(SSB)，在氨基酸序列和DNA结合性质上都与大肠杆菌SSB相似。与在mtNDA复制的作用一致，多种来源的SSB可体外促进聚合酶γ活性，大鼠的SSB可

结合在D环区域的亲代H链上。在线粒体中也普遍发现了DNA螺旋酶的存在。

二、滚环复制

有一种单向复制的特殊方式称为滚环复制(rolling circle replication)。在滚环复制中，只有一条链复制而产生一些环状分子的拷贝。其中一条链发生断裂，产生了自由的3′羟基末端。然后以此为引物，以未断裂的一条环状链为模板，由DNA聚合酶催化延伸，产生叫做滚环的结构。复制反应可看成生长点沿着环状的模板链滚动。随着它的滚动，3′端不断延长，5′端不断地被置换甩出而成为一条单链。随着反应的循环进行，在单链上合成了环状基因组的许多单位的拷贝，这种单链尾巴按照单位长度断裂就会产生原初环状复制子的单链线性拷贝。这个线性DNA可保持其单链形式存在或者通过合成其互补链而转化为双螺旋形式。因此滚环复制提供了扩增原初复制子的方法(图4-25)。

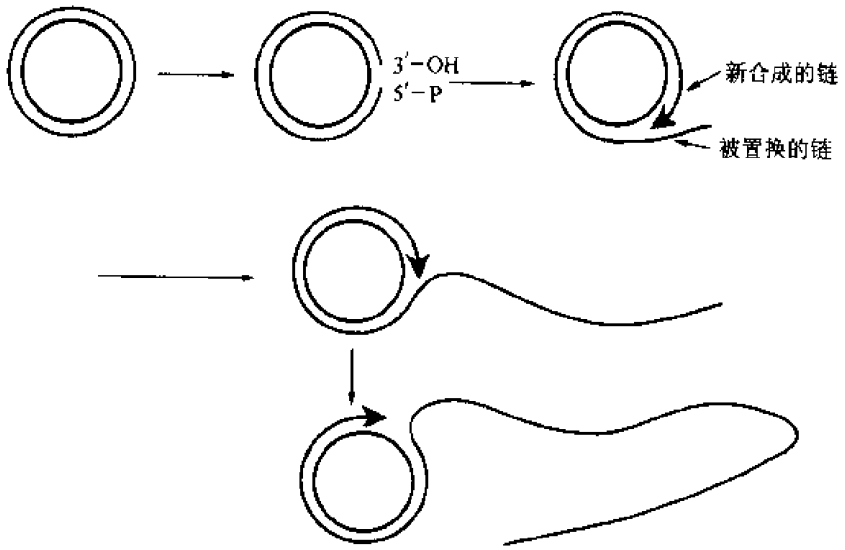


图4-25 滚环复制

在噬菌体中滚环复制是一个普遍现象。噬菌体 $\Phi X174$ 是单链环状DNA分子(正链)，它在复制过程中首先合成其互补的负链，形成共价闭合的双链超螺旋分子，然后进行滚环复制合成大量正链。噬菌体基因组编码产生A蛋白，它识别双螺旋DNA正链上的特殊位点(复制起点)，并在其上产生一个切口(nick)，游离出一个3′-OH和一个5′-磷酸末端，并与其5′末端共价结合。随后在DNA聚合酶催化下，以环状负链为模板，从正链的3′ OH末端加入脱氧核苷酸。新链沿着环状的负链模板不断延长，而结合着A蛋白的5′端不断被置换出来。SSB蛋白结合到甩出的单链上，使其环化回折。所以当新链的生长点重新返回复制起点时，被取代的正链5′末端的A蛋白也在起点附近，它将再次识别起点并将链切断。被取代的正链作为一个环被释放。A蛋白参与了环化过程，使正链产物的3′和5′末端连接在一起。然后A蛋白又重新连接于双链环正链切口处的5′磷酸

末端，开始下一次循环。通过多次滚环复制，可产生许多拷贝的环状正链。这些正链可用作模板合成互补的负链，以形成双链环状DNA，重新参加滚环复制，也可以包装到噬菌体的蛋白质颗粒中以产生后代噬菌体。

实验证明，某些双链DNA的合成也可以通过滚环复制的方式进行。例如λ噬菌体复制的后期，以及非洲爪蟾的卵母细胞中rRNA基因的扩增都是以这种方式进行的。

滚环和D环的存在表明，可以只以一条链为模板起始DNA的合成。双螺旋的开放结构并不会引起另一条链起始合成，两条链的复制起点可以分别在不同的位置。

第八节　真核生物的复制过程

一、概　　况

在真核生物细胞的细胞周期中，DNA的复制发生在中期的S期，高等真核生物细胞的复制常常持续几个小时。真核染色体中DNA复制是通过许多独立的复制子来完成的，其中只有一部分复制子在S期的任何时间都参与复制，而其他复制子只在S期的特定时间活化。每个复制子都有固定的起点（复制起点），常常采取双向复制方式。

与原核生物不同，真核生物的复制子相对小，而且它们的复制叉移动的速度比细菌复制叉的运动要慢许多，复制速度大约500～5000 bp/min。如果所有的复制子同时发挥作用，哺乳动物基因组的复制将在一小时内完成，但在典型的体细胞中，S期可持续6小时，表明在任何时间内，活化的复制子不会超过15%。

原核生物细胞的基因组只有一个复制子，就是细菌染色体本身，所以复制和分裂的单位是一致的。真核生物染色体是线性双链分子，含有许多复制起点，因此是多复制子（multireplicon）。所以一个分裂单位包括很多复制单位，这使得复制的控制很复杂。在一个细胞周期中，染色体上所有的复制子都起始复制，尽管它们不是同时活化的，但都是在一个周期内被激活的，而且每个复制子在一个细胞周期只能被激活一次，这表明尚未复制的复制子上一定有某些信号，可被复制体系鉴定出来并在其上开始复制，结果一个复制子在细胞周期内不能复制两次。而且由于许多复制子是独立活化的，必然存在另一些信号表明基因组中所有复制子的复制过程已全部结束，这种调控机制尚不清楚。

有证据表明真核生物染色体的复制子中没有复制终止，复制叉从其起点连续移动直到遇到邻近复制子的复制叉时终止复制。

二、DNA聚合酶

与真核细胞中多起点复制相应的特点是细胞中DNA聚合酶的数量众多，在大肠杆菌中，只有10～20个PolIII全酶，然而一个典型的动物细胞中含有20 000～60 000分子的聚合酶α。

在哺乳动物细胞抽提物中发现了五种不同的DNA聚合酶，它们是α、β、γ、δ、ε。α、β、δ、ε位于核中，而γ是线粒体酶。其中3个核酸酶α，δ，ε属于α样的DNA聚合酶，有共同的结构和催化特征。

表 4-6 真核生物 DNA 聚合酶

DNA 聚合酶	α(I)	δ(III)	ε(II)	β	γ
位置	核	核	核	核	线粒体
合成功能	引发	DNA 合成	DNA 合成 修复	修复	复制
其他功能		3′-5′ 外切核酸酶	3′-5′ 外切核酸酶		3′-5′ 外切核酸酶
抑制剂	aphidicolin	aphidicolin	aphidicolin	双脱氧 TTP	双脱氧 TTP

这些酶中哪个是真正的复制酶呢？利用 SV40 DNA 的复制作为模型研究哺乳动物 DNA 复制，发现在体外 α 和 δ 两个酶就足以复制病毒的环状基因组，在体内结果也一样。DNA 聚合酶 α 是一个多亚基复合物，由 4 个不同的亚基组成，大亚基(180 kDa)具有聚合酶活性位点。两个小亚基(50～60 kDa)具有引发酶活性，能合成短的 RNA 引物。哺乳动物的 DNA 聚合酶 α 进行性不高，每次结合事件合成核苷酸少于 100 个。DNA 聚合酶 δ 与 DNA 聚合酶 α 不同，它含有 3′,5′外切核酸酶活性，而缺少内在的引发酶活性。DNA 聚合酶 δ 还没有被很好地鉴定。DNA 聚合酶 α 和 δ 在酵母中的同源物分别叫做 DNA 聚合酶 I 和 DNA 聚合酶 III，它们负责酵母核的复制。无论在体内还是在体外复制，DNA 聚合酶 α 和 δ 对抑制剂 2′,3′-双脱氧胸腺嘧啶(dideoxythymidine)和抗菌素蚜栖菌素有相同的反应，这些结果使人们认为它们都是核 DNA 的复制酶，人们还认为 DNA 聚合酶 α 合成后随链而 DNA 聚合酶 δ 合成前导链，但现在看来 DNA 聚合酶 α 的功能是起始复制，而 DNA 聚合酶 δ 的功能是延伸两条链，因此前导链和后随链冈崎片段的起始具有相同的程序。但目前我们还不知道前导链和后随链合成的起始是否是由同一个 DNA 聚合酶 α/引发酶复合物完成的。

聚合酶 ε 似乎不参与 SV40 DNA 的复制，但最近研究发现 ε 催化亚基的表达与人细胞增殖紧密联系，表明 ε 在细胞 DNA 复制中起关键作用，至少与复制紧密相关。此外，DNA 聚合酶 ε 也可能参与修复反应，如人成纤维细胞的 UV 损伤修复、重组引起的双链断裂和缺失的修复、精子发生过程中减数分裂时的重组和修复。人 HeLa 细胞的 DNA 聚合酶 ε 有两个亚基，分别为 >200 kDa 和 55 kDa，酵母聚合酶 ε 有 5 个亚基，分别 > 200 kDa、80 kDa、34 kDa、31 kDa、29 kDa。

DNA 聚合酶 β 的作用是碱基切除修复，而 DNA 聚合酶 γ 是线粒体的复制酶。DNA 聚合酶 γ、聚合酶 δ 和聚合酶 ε 都有 3′→5′核酸外切酶的校正活性。

DNA 聚合酶 δ 的进行性由 PCNA 滑动钳蛋白来保持，作为 DNA 聚合酶的进行性因子，协调两个聚合酶的行为。PCNA 的晶体结构与大肠杆菌 β 亚基十分类似，PCNA 三聚体形成环围绕着 DNA。尽管其序列和亚基组成与二聚体的 β 钳有所不同，但功能是相同的。

目前已经纯化了体外复制真核生物 DNA 的系统。它们的功能与原核生物中的类似，但在使用上有所不同，见表 4-7 。

表 4 - 7 不同生物的复制体系

功能	大肠杆菌	噬菌体 λ	噬菌体 T4	SV40/人	酵母
起始蛋白	DnaA	λ的O蛋白		T抗原	ORC
装载和改型因子	DnaC	λ的P蛋白，DnaJ，DnaK	gp59	细胞伴侣?	Cdc6蛋白
螺旋酶	DnaB		gp41	T抗原(SV40)	MCM蛋白?
引发酶	DnaG引发酶		gp61	Polα-引发酶	Polα-引发酶
聚合酶	PolIII全酶的α亚基		gp43	Polδ	Polδ和Polε
校正作用的外切核酸酶	PolIII全酶的ε亚基		gp43	Polδ	Polδ和Polε
滑动钳	β亚基		gp45	PCNA	PCNA
滑动钳载体	γ复合物		gp44/62	RF-C	RF-C
单链DNA结合蛋白	SSB		gp32	RP-A	RP-A

三、SV40 DNA 复制

SV40病毒(simian virus 40)只编码一个蛋白质(T抗原)以供基因组复制的需要，其他复制所需的蛋白质由宿主细胞提供，SV40 DNA复制发生在宿主细胞核中，在病毒和细胞DNA复制之间有很多相似之处，这使SV40病毒成为研究真核生物细胞DNA复制的理想模型体系。

1. DNA 复制起点

SV40病毒基因组由大约5000个碱基对的环状双螺旋DNA分子组成，病毒基因组与组蛋白复合形成与细胞染色质类似的核蛋白结构。在SV40复制起点处有两个T抗原结合位点(图4-26)其中核心起点(位点II)是基因组中一个64 bp的片段，它含有在体内和体外起始病毒DNA复制所需的所有核苷酸序列元件。遗传学分析显示核心起点组成是复杂的，至少由三个功能不同的序列微区组成。核心起点中心是27 bp的严格的反向重复序列，其中含有4个拷贝的五核苷酸序列特征微区(GAGGC)，这些序列是病毒起始蛋白T抗原的识别位点。在T抗原结合位点的一侧是一个由A/T组成的17 bp片段，可能是起始解链的位点。T抗原结合位点另一侧是15 bp不严格的回文结构，其功能未知。

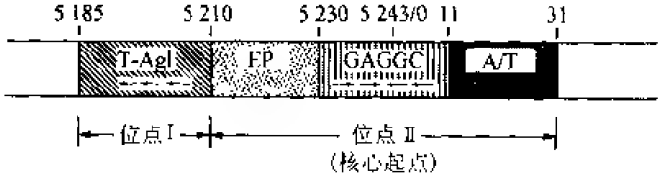


图 4 - 26 SV40 DNA 复制起点

后两个侧翼区域在起始 SV40 DNA 复制时将经历结构转化。SV40 DNA 复制需要起点具有所有三个微区，而且它们之间间隔距离对起点功能也很关键。尽管 64 bp 核心起点区足以支持 SV40 DNA 复制起始，核心区外的序列却能影响起始的效率。第二个 T 抗原结合位点(位点 I)位于核心起点的附近，可使复制效率提高 1～4 倍。与 SV40 DNA 转录活化相关的元件，如 SV40 增强子或转录因子 SP－1 的结合位点，对起始复制也很重要，在核心起点附近存在这些序列元件可至少使 DNA 体内复制效率提高 10 倍。

2.SV40 的大 T 抗原

在 SV40 起点起始复制需要病毒基因产物 T 抗原结合起点并起始链分离过程。SV40 的大 T 抗原(SV40 large tumor antigen)是病毒编码的含磷蛋白质，含有一个分子量 82 kDa 的多肽。这个蛋白质在病毒 DNA 复制起始过程中起起始因子、DNA 螺旋酶和引发体装载蛋白的三重作用。

从 SV40 *ori* 起始的双向 DNA 复制经历以下步骤：在 ATP 的促进下，T 抗原会在起点聚集形成六聚体，两个 T 抗原六聚体识别并结合到 *ori* 上。结合研究表明 T 抗原分子与起点中所有 4 个五核苷酸重复结合，形成核蛋白结构。T 抗原的结合在起点 DNA 上产生了显著的结构变化，使 15 bp 不完全回文结构处的双螺旋去稳定，并在富含 AT 区域造成结构变形。在解旋过程中 T 抗原与富含 AT 区域产生序列专一性的结合。(图 4－27)通过对富含 AT 片段断的碱基取代突变分析表明解旋效率不仅取决于其碱基组成，而且取决于其精确核苷酸序列。T 抗原有明显的螺旋酶活性，一旦结合起点，它能够进入双螺旋中并催化依赖 ATP 的双链解旋。解旋是建立复制叉和产生引发和延伸新链的底物的关键步骤。T 抗原的螺旋酶活性也参与延伸过程。

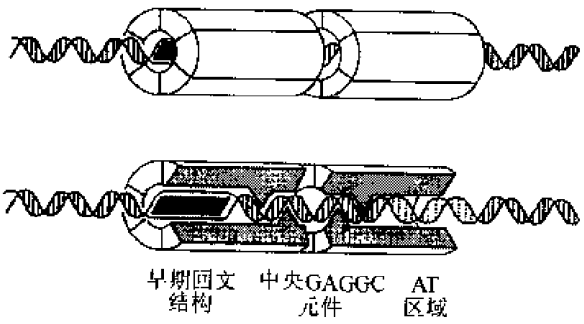


图 4－27　SV40 的大 T 抗原

除了专一性的核苷酸序列元件外，T 抗原介导的解旋反应还需要宿主细胞提供辅助蛋白，如需要单链 DNA 结合蛋白防止解旋时暴露的单链重新结合。SV40 的单链结合蛋白是复制蛋白 A(replication protein A)也叫做 RP-A、RF-A 或真核生物 SSB，它由三个紧密结合的亚基组成，分子量分别为 70 kDa、32 kDa、14 kDa，大亚基可专一性结合单链 DNA。起始蛋白 T 抗原－DNA 复合物结合 RPA，增强了 SV40 T 抗原的 DNA 螺旋酶活性，催化进一步解旋。在解旋反应中，大肠杆菌的 SSB 可取代 RP-A 作为单链结合蛋白，但是它不能代替 RP-A 完成复制反应，表明 RP-A 在复制中还有其他作用。

T 抗原和大肠杆菌 DnaA 蛋白一样，都能与单链区域的一条链发生专一性相互作用。因此除了专一性识别 DNA 结合位点外，T 抗原在起点解旋过程中也可以作为单链 DNA 结合蛋白发挥作用。

3. RNA 引物和起始 DNA 的合成

T 抗原具有引发体装载功能，并可激活 Polα 的引发酶活性。在起始的下一步，双功能的 DNA 聚合酶 α/引发酶识别并结合 T 抗原-RPA 起点复合物，从而导致了引发体的形成，开始引物合成。引发反应是异常的：首先利用 Pol α 的引发酶活性合成一个短的 RNA 引物，然后立刻用同一个 Pol α/引发酶的 DNA 聚合酶活性快速延伸所形成的 RNA，合成短的起始 DNA(initiator DNA，iDNA)。iDNA 长度通常为 3～4 nt，但在纯化的 DNA 复制体系中，Pol α/引发酶合成的 RNA-DNA 产物要长得多(大约 300nt)，表明有其他因子调控了 Pol α 的活性。这个片段长度也受 Polα 进行性的限制。

当引物合成并起始了 DNA 合成后，复制叉形成并开始复制的延伸期。SV40 复制叉上至少包含了两种 DNA 聚合酶(DNA 聚合酶 α/引发酶复合物、DNA 聚合酶 δ、还可能有 DNA 聚合酶 ε)、聚合酶辅助蛋白(PCNA 和 RFC)，延伸也需要 RPA 的参与，它可促进 DNA 聚合酶的活性。除了在 DNA 合成起始前解旋起点外，T 抗原也是延伸过程中负责亲代 DNA 解旋的螺旋酶。在存在单链结合蛋白的情况下，T 抗原内在的螺旋酶活性可以高度进行性的方式解旋长的双螺旋 DNA 片段。研究表明 T 抗原在 DNA 单链上以 3′→5′方向运动。因此如果 T 抗原是负责复制叉移动的螺旋酶，它应该沿前导链运动。相反所有已知直接参与原核生物复制的螺旋酶都沿后随链运动。SV40 DNA 复制需要拓扑异构酶活性，在复制过程中拓扑异构酶起两个作用，一是释放复制叉前进产生的超螺旋张力，以保证亲代链解旋。两个已知的哺乳动物 DNA 拓扑异构酶，拓扑异构酶 I 或 II，都有这个功能。第二个作用是在复制完成时介导新合成的子代双螺旋分离。

4. 从 Pol α 到 Pol δ 的转换

起始前导链 DNA 复制需要从 Pol α 到 Pol δ 的转换。一旦 iDNA 合成，细胞复制因子 C(RFC)结合到 iDNA 的 3′末端，在模板上装载 PCNA 和 DNA 聚合酶 δ，取代聚合酶 α。因此 Pol α/引发酶不能一直完成 DNA 合成。RFC 是一个多亚基蛋白质，它的三个亚基分子量分别为 140 kDa、41 kDa、37 kDa，它的作用是真核生物的滑动钳载体。其中 140 kDa 的大亚基识别并结合引物-模板连接，而 41 kDa 的亚基具有依赖 DNA 的 ATP 酶活性。

PCNA 是 DNA 聚合酶 δ 的进行性因子，其单体分子量为 37 kDa，这个蛋白质与人们先前发现的增殖性细胞核抗原(proliferating cell nuclear antigen，PCNA)或细胞周期蛋白(cyclin)是同一个蛋白质。PCNA 的晶体结构与大肠杆菌 β 亚基十分类似，它是一个环状的三聚体，具有 3.5nm 的洞，足以容纳双螺旋 DNA 在其间滑动而不与蛋白质产生物理接触。尽管其序列和亚基组成与二聚体的 β 钳有所不同，但功能是相同的。PCNA 参与延伸反应，可提高 SV40 DNA 复制效率，在没有 PCNA 时，DNA 合成也能起始，但只合成很短的新链，最多几百个核苷酸；而存在 PCNA 时，DNA 聚合酶 δ 进行性提高，每次结合事件可催化聚合至少 1000 个核苷酸，但 PCNA 对 DNA 聚合酶 α 的活性和进行性没有作用。进行性的 RFC/PCNA/Pol δ 复合物延伸新的 DNA 链，形成连续合成的前导链复合

物。

目前认为后随链 DNA 复制是以冈崎片段形式半不连续进行的。与大肠杆菌或 T4 噬菌体二聚体 DNA 聚合酶机制一致，双链合成是一个高度协同的过程。复制叉处含有一个 DNA 聚合酶 α/引发酶复合物和两个 DNA 聚合酶 δ，Pol α/引发酶在由 RPA 包被的后随链模板上不断引发新冈崎片段，然后利用与前导链起始过程中相同的机制实现 Pol α 到 Pol δ 的交换。两个 DNA 聚合酶 δ 一个合成前导链，而另一个合成后随链的冈崎片段。后随链产生环回折使后随链/前导链聚合酶复合物能够一起沿 DNA 移动。与 DNA 聚合酶 α 不同，DNA 聚合酶 δ 含有起校正作用的 3′→5′外切核酸酶活性，可保证双链 DNA 的精确合成。

5. 冈崎片段连接

在冈崎片段合成后，RNA 酶 H1 和 5′→3′外切核酸酶 MF1 共同作用切除 RNA 引物，后随链的 DNA 聚合酶 δ 填充两个冈崎片段之间的缺口，DNA 连接酶 I 用来填补完成的冈崎片段之间的切口，产生完整的后随链。

四、酵母复制

通过对 SV40 DNA 复制的研究，人们对真核生物的复制体系和复制过程有了一定的认识，但 SV40 毕竟不是真正的真核生物，因此其复制过程、复制调控与真核生物复制的实际情况相比，仍有一些差别。近年来，人们通过研究低等真核生物酵母的复制，进一步了解了真核生物的复制情况。

1. 酵母的复制区和复制起点

对大多数真核生物 DNA 复制起点的性质人们知之甚少，许多知识是通过低等真核生物酵母获得的。人们从酿酒酵母（*Saccharomyces cerevisiae*）染色体上克隆了短的 DNA 序列，含有这些序列的质粒可在细胞循环 S 期进行自主复制。表明这些专一性 DNA 序列能指导 DNA 复制的起始，被命名为自主复制序列（autonomously replicating sequence，*ARS*）。*ARS* 序列在质粒和酵母正常染色体复制中都发挥 DNA 复制起点的功能。

在酵母染色体中，只有部分 *ARS* 序列可作为复制起点，即染色体中 *ARS* 元件比染色体中活化的复制起点多。即使这样，在酵母染色体中复制起点也是过量的，即在染色体中潜在的复制起点比真正需要的复制起点多，可能在细胞分化中允许选择性地使用不同的复制起点。

人们了解最清楚的 *ARS* 序列是位于酵母染色体 IV 着丝粒附近的 *ARS1* 序列，已经确定它是酵母染色体的一个复制起点。*ARS1* 可被分为三个功能区，*A*、*B*、*C*。*A* 和 *B* 已足够完成 *ARS* 的功能，构建有效的复制起点。所有其他 *ARS* 序列都有相同的 *A-B* 组织结构，而 *C* 区可能只存在于 *AR1* 上，*C* 区只有微弱的起始活性，而且只有在缺少 *B* 区时 *C* 区才能发挥作用。

ARS1 的 *A* 区是 15 bp 的 DNA 序列，其中含有 11 bp[5′-(T/A)TTTA(T/C)(A/G)TTT(T/A)-3′]的保守序列，它是所有酵母 *ARS* 中都具有的一致序列，叫做 *ARS* 共有

序列(*ARS* consensus sequence, *ACS*)。这个序列对 *ARS1* 的功能十分重要。当 *A* 区中只有 *ACS* 序列时,起点的起始功能将下降 20 倍,因此 *ACS* 以外的序列对 *A* 区的功能也很重要。

对 *ARS1* 中的 *B* 区已进行了详细的研究(图 4-28),它位于 *ACS* 序列富含 T 的 DNA 链的 3′一侧,是大约 80 bp 的一个开阔区域。*B* 元件含有在超螺旋质粒中很容易解旋的 DNA 序列,最近的研究表明 *B* 元件作为一个复制区元件,其作用与 *A* 元件同等重要,它可能促进复制反应的起始。

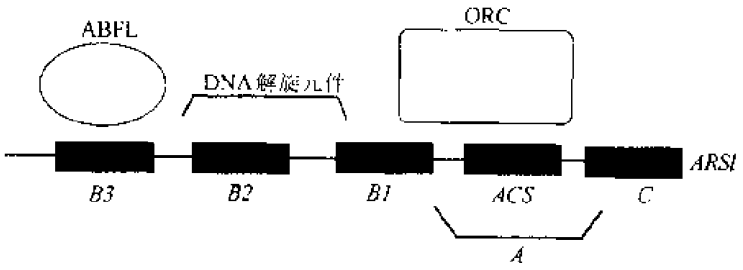


图 4-28 酵母的复制区和复制起点

B 区包括三个亚功能区,*B1*、*B2*、*B3*。*B1* 和 *B2* 的功能未知,在 *B2* 的 11 bp 中有 9 bp 与 *ACS* 序列一致,而 *B1* 是新的序列。它们可能结合 DNA 复制蛋白或容易产生一种不稳定结构而作为 DNA 解旋位点,也许两者都有。当 A 区存在时,在 B 的三个亚区中只要存在任意两个区域都足以保持有效的起点功能。任何一个区域单独存在,即使在多拷贝的情况下,也不能实现 *ARS* 的作用,因此三个亚功能区的功能不同,不能互相取代。

ARS1 的 *B3* 元件是 *ARS* 结合因子 1(*ARS*-binding factor 1)的结合部位。ABF1 是一个多功能的、序列专一性的 DNA 结合蛋白,作为转录因子参与转录活化和转录沉默。此外,它也是 DNA 复制的辅助因子,它的结合是复制所需。因此与真核病毒复制一样,在真核染色体中也存在复制起始和转录起始的直接联系。作为 ABF1 的结合位点,*B3* 序列也可被其他转录因子的结合位点取代。

总之,真核生物的染色体复制起点由一些 DNA 元件以一定的排列形式组织而成,其中 A 区具有复制起始子的功能,在构成 *ARS* 功能的序列中,只有 *A* 区是必须的,而 *B1*、*B2*、*B3* 也比较重要,4 个元件足以构成有效的起点。在 DNA 复制的 *ARS* 起点至少有两个活化位点,一个是 ABF1 结合位点,既能活化复制也能活化转录;另一个是复制真正的活化位点。

在酵母基因组中与 A 区一致的序列很多,可以使质粒高频转化的染色体序列中只有一小部分是其染色体上真正的复制起点。这说明只有 A 区还不够,A 区还需要与多个具有活化作用的元件结合,才能成为有活性的复制起点,如同 *ARS1* 中的情况一样。

2. 真核生物的起点识别复合物

酵母的起始因子是一个叫做起点识别复合物(origin recognition complex, ORC)的多亚基蛋白质,它由 6 个不同的多肽组成,分子量分别为 120、72、62、56、53、50 kDa,这 6 个亚基对细胞分裂和 DNA 复制起始都有重要作用。遗传分析表明 72 kDa 的 ORC2 基因突

变可导致 DNA 复制失败。与野生型细胞相比，ORC 温度敏感型突变体以很高的速率失去质粒，同时在非允许温度下细胞不能进入 S 期，表明 ORC 在 DNA 复制起始时的正调控作用。

ORC 可在溶液中预先组装，并在 ATP 的作用下结合起点 DNA。DNA 结合实验表明 ORC 可与 *ARS1* 复制区关键的 *A* 元件结合，ORC 与所有 *ARS* 元件结合都需要识别 *ACS* 序列，但对 B 的各个亚区没有序列上的特殊要求。除了结合 *A* 区外，ORC 还能与 *B1* 相互作用。结合 ORC 的 *B* 元件区域在 DNase I 作用下形成了周期性的酶切图谱，表明 DNA 缠绕在 ORC 多肽上，就像组蛋白八聚体在核小体中缠绕 DNA 那样，不同的是，ORC 结合 DNA 需要依赖 ATP 的存在。

可见 ORC 具有原核生物 Dna A 起始蛋白的一些特性，表明它可能是真核细胞复制的起始蛋白。此外，ORC 可作为单链结合蛋白，选择性地与双链 DNA 结合位点处的一条链相互作用，但没有观察到 ORC 促进的 DNA 解旋。

从不同生长时期细胞中提取染色质，其核酸酶切图谱表明 ORC 几乎在整个细胞周期中一直结合 *ACS* 序列，这表明尽管 ORC 对 DNA 复制是重要的，但是仅仅结合 ORC 不足以决定 DNA 合成起始。因此可能有与细胞循环相关的其他因子，与 ORC 相互作用以控制 S 期 DNA 合成起始，而 ORC 作为这些因子结合 DNA 的一个装载平台（如 λ 噬菌体的 O 蛋白一样），成为复制调控的一个关键因素。通过对 ORC 蛋白亚基序列分析无法推测其功能，这是因为其中只有 ORC1 与其他蛋白质有同源性。ORC 与出芽酵母的 Sir3 蛋白同源，Sir3 在转录沉默中发挥重要作用，它与 ORC1 N 端的 200 个氨基酸有 50% 同源性。ORC1 还与参与细胞分裂循环的蛋白质同源：它们是酿酒酵母（*S. cerevisiae*）的 Cdc6 蛋白和裂殖酵母（*Schizosaccharomyces pombe*）的 Cdc18 蛋白，这两个蛋白质在 DNA 合成调控中起关键作用。ORC1、Cdc6、Cdc18 的同源序列包括一个 NTP 结合区，ORC5 也含有一个 NTP 结合区。如上所述，ORC 结合 *ACS* 需要 ATP，突变分析表明 ORC1 的 NTP 结合区对细胞生存是关键的，而 ORC5 的 NTP 结合区突变将导致细胞生长对温度敏感。ORC 中存在两个核苷酸结合位点，其中只有一个对 ORC 功能关键，表明 ORC 除了作为复合物装载平台外，还具有其他功能。

ORC 不仅存在于酵母中，最近在许多物种甚至人中都克隆了 ORC 亚基 1、2、5 的同源物，表明酵母中调控 DNA 合成的组分和机制在所有真核生物中都是保守的。

3. 真核生物的装载因子和螺旋酶

目前尚未确切发现负责真核生物复制的六聚体螺旋酶，最有可能成为真核生物复制性螺旋酶的候选者是 MCM 蛋白复合物。它是依赖 DNA 的 ATP 酶家族的成员，最早作为在酵母中保持小染色体的基因而被鉴定。最近的研究显示一种无进行性的 DNA 螺旋酶活性可与一系列人的 MCM 蛋白共同纯化。另外，对酿酒酵母的体内观察表明 MCM 蛋白定位于起点而且它们随着复制叉移动。尽管有这些进展，还需要进一步的研究以确定观察到的 DNA 螺旋酶活性是否是由于 MCM 蛋白的直接作用，以及这些蛋白质是否一直存在于复制叉上。

尽管还不能确定真核细胞中复制性的螺旋酶，但最近通过研究细胞起点处蛋白质组装表明，真核生物有与原核生物相似的复制叉组装过程。对爪蟾抽提物的体外复制研究

和对酿酒酵母的体内复制研究表明 Cdc6 蛋白(可与 ORC 相互作用,为细胞进入 S 期所必需)和 MCM 蛋白在 DNA 合成起始前以一定的顺序组装到 ORC-起点复合物上。另一个可以与 ORC 和 MCM 蛋白相互作用的蛋白质,Cdc45 蛋白,也在起始前结合到起点上。有证据表明 MCM 蛋白作用于复制叉上,而且还发现 Cdc6 蛋白是起始过程所必需而不是延伸所需要,表明 Cdc6 蛋白与 DnaC 和 λ 噬菌体的 P 蛋白类似,是一个装载因子,而 MCM 蛋白与 DnaB 一样有螺旋酶的功能。Cdc45 的功能还不清楚,它能随复制叉移动,它的功能可能是协调真核生物聚合酶组装和功能的发挥,类似大肠杆菌的 τ 蛋白。

4. 复制过程

对真核生物起点结合复合物的研究表明,伴随复制起始也有类似的改型事件。在起始过程中,与起点结合的蛋白质组分明显地改变。这些改变包括释放或降解装载因子 Cdc6 蛋白,以及最后起点结合复合物的去组装。控制这些变化的机制还不清楚,但有一些可能性。细胞循环控制的激酶对复制前复合物(在 G_1 期位于真核生物起点的蛋白质聚集体)的一个或多个组分的修饰可能触发改型。另外,ORC 和 Cdc6 蛋白对 ATP 的水解也可能触发起点处复合物的蛋白质组成和构型的变化。在 G_1/S 转变时 Cdc6 蛋白的降解是起始复制前改型组装蛋白的另一个机制。

在起点结合蛋白和聚合酶之间的其他相互作用也可能在起始时发挥功能。例如,真核生物的起始蛋白(也是螺旋酶)直接与 DNA Pol α/引发酶作用并使其组装到起点。

5. DNA 聚合酶 ε

目前还不清楚两个真核生物聚合酶 Pol δ 和 Pol ε 的具体功能,在不同的复制子中两个酶的作用是不同的。SV40 DNA 可以在无 Pol ε 的情况下复制,而对酵母的遗传分析表明 DNA 聚合酶 ε 是细胞生存和完成 S 期所必需的。可能在复制相对短的 SV40 复制子时 Pol ε 是可有可无的,但在复制真核生物染色体上相对长的复制子时仍需要 Pol ε。在染色体复制时,两个酶分别合成前导链和后随链,但还不能确定具体哪个聚合酶作用于哪条链。在人和酵母中,Pol ε 还参与 DNA 修复。

6. 5′→3′外切核酸酶活性

真核生物的 DNA 聚合酶没有 5′→3′外切核酸酶活性,切除 RNA 引物的过程由一个独立的叫做 FEN-1 的外切/内切核酸酶完成。这个酶专一性作用于双螺旋 DNA 的未退火的 5′单链尾巴。当完成了冈崎片段的合成后,螺旋酶置换含有 RNA 引物链的 5′末端产生了 FEN-1 的底物。已知在酿酒酵母中这个螺旋酶是 Dna2 蛋白,这个蛋白质可与 FEN-1 发生相互作用。RNaseH1 也可能参与引物切除。PCNA 与 FEN-1 相互作用使 FEN-1 组装到其作用位点。一旦引物被切除后,DNA 聚合酶填充沟,留下了 DNA 连接酶连接的缺口。

五、真核生物复制时的核小体结构

染色体是真核生物 DNA 的存在形式,它的复制要经历另外的解离步骤(DNA 和组蛋

白八聚体的解离)和组蛋白-DNA的重结合步骤。DNA是半保留复制的,而组蛋白八聚体是以全保留方式传给子代分子的,即亲本的组蛋白八聚体在复制过程中并不解离,而新组蛋白的8个亚基则完全是重新合成的。而且在复制过程中亲代八聚体仍与亲链保持一定的结合,并不完全从亲本链上解离下来,从而导致复制后所有的亲代八聚体都分布在一条子链上,而新的八聚体都分布在另一条子链上。在SV40的复制体系中,亲代八聚体都结合在前导链上。

六、真核生物染色体末端的复制与端粒酶

真核生物同样面临线性DNA复制时所产生的5′末端隐缩的问题,常规的真核DNA聚合酶同样不能完全复制真核生物的染色体末端(端粒)。问题的解决需要一种特殊的聚合酶——端粒酶(telomerase),这种酶可以保持染色体端粒的长度。

在70年代以前,端粒的精确组成尚未确定,1978年Blackburn等人发现一种具有纤毛的单细胞生物四膜虫(*Tetrahymena*)的端粒含有一种短的、简单的核苷酸序列重复(TTGGGG)。此后,科学家们考察了包括动物、植物和微生物在内的许多生物中端粒的特征。实际上所有的端粒-包括鼠、人和其他脊椎动物的重复单位中常含有较多的T和G核苷酸。例如人和小鼠的端粒以TTAGGG序列为特征,而线虫的端粒则具有TTAGGC的特征。

80年代初期的研究揭示,由于某种原因,不同的有机体之间端粒重复亚单位数目不同,甚至在同一有机体的不同细胞之间也不同。此外随着时间的推移,特定细胞中的重复亚单位数目也是波动的。但是每个物种有一个特有的平均数。在四膜虫中一般的端粒有70个重复;而人的端粒有2000个重复。所观察到的不均一性使Blackburn等人提出了末端复制问题的新的解决办法。

Blackburn等认为细胞在端粒长度上所观察到的波动是细胞企图将端粒保持在一个大致恒定的尺寸上的信号。在细胞分裂过程中端粒缩短,但它们也由于结合上新合成的端粒亚单位而被延长。人们推测这些亚单位是由一种特殊的DNA聚合酶合成的,它不同于以往的正常的DNA聚合酶。研究人员推测这种特殊的酶能够在没有现成DNA模板可利用的情况下从缩短处重建DNA单链。1984年,Blackburn等发现了这种特殊酶(端粒酶)的存在,他们把合成的端粒与四膜虫的细胞抽提物混合在一起,结果发现端粒得到了延伸。随后人们对端粒酶的性质进行了详细的研究,发现这种酶是一个核糖核蛋白,既含有蛋白质成分也含有RNA分子,在RNA上含有复制端粒亚单位所需要的关键核苷酸模板。因此端粒酶可以看作一种特殊的DNA聚合酶,即自身携带RNA模板的反转录酶。(图4-29)此后不同的研究人员在酵母、蛙和小鼠中发现了端粒酶。1989年,Morin首次在人的癌细胞系中发现了端粒酶。该酶的精确组成在不同物种之间有所不同。

在许多单细胞有机体中端粒酶的重要性是无可争辩的。例如四膜虫需要端粒酶以保持其永生性,如果端粒酶发生了改变,端粒就会缩短,细胞也将死亡。在酵母中端粒酶也发挥同样的作用。但在多细胞有机体甚至在人体中端粒酶是否也有同样的功能呢?

令人惊奇的是人类的许多细胞中都缺乏端粒酶,大多数普通的体细胞在培养中分裂时都失去了其端粒片段,而且在正常人体组织中的端粒随着人的衰老也发生缩短。而在

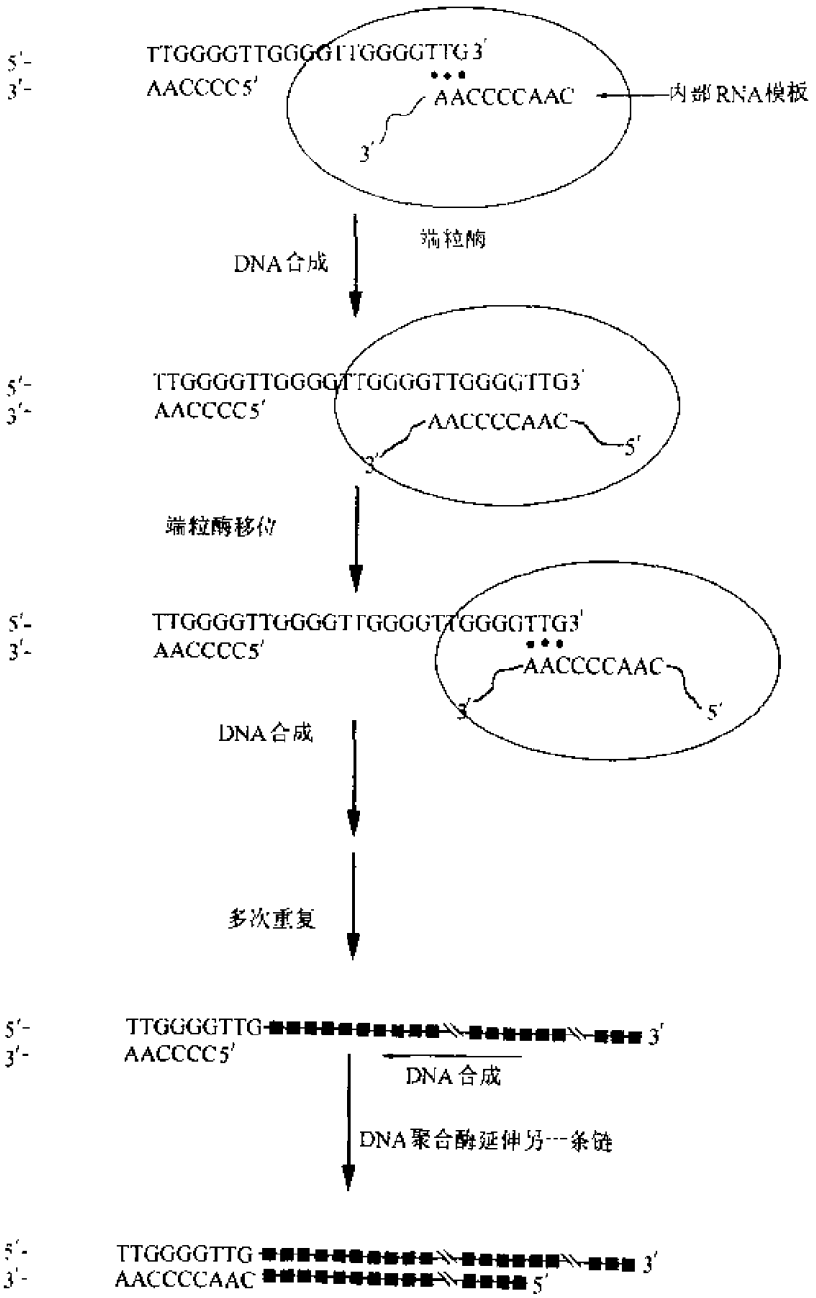


图 4－29　端粒酶

种系细胞中端粒可被完整保存下来。这些结果表明人体细胞能通过监测失去的端粒重复数而计数分裂次数，当端粒长度下降到某一临界值时它们就可能终止分裂。随着时间的推移，端粒的缩短和细胞增殖能力的下降也许是人衰老的原因之一。在最近的实验中，研究人员通过在正常的体细胞中引入端粒酶催化亚基，使细胞额外地多进行了 20 次分裂。

另外，人们也推测端粒酶与肿瘤的关系。在理论上，体细胞缺乏端粒酶，通过连续分裂会失去它们的端粒，这种端粒缩短被细胞的特定机制监测，从而在没有造成大的损害前

终止细胞的分裂。如果肿瘤细胞产生端粒酶，那么这些细胞将保持它们的端粒而有可能无限地分裂并生存下去。

人们发现，端粒酶不仅在癌细胞系中是有活性的，而且在大多数人的肿瘤样品中也都有活性，在正常的人体组织中几乎不能检测到其活性。研究人员还发现，人肿瘤中的端粒比周围正常组织中的端粒短得多。

这些结果暗示癌细胞中的端粒很短，因为只有在其已开始无控制的复制之后细胞才合成端粒酶。此时细胞已失去很大数量的端粒亚单位。当端粒酶最终被活化时，它将稳定这严重缩短的端粒，使过度增殖的肿瘤细胞变成不死的细胞。根据这些结果，人们提出了端粒酶被人体正常和恶性细胞活化的方式。根据这种模式，端粒酶通常是由发育胚胎中种系细胞产生的，但是一旦人体完全形成，大多数体细胞中的端粒酶就会被抑制，同时随着细胞增殖，端粒缩短。当端粒下降到一个阈值水平，将发出防止细胞进一步分裂的信号。但是，如果导致癌的遗传突变阻止了这种信号的正常传递，细胞将越过正常的衰老并继续分裂，细胞将失去端粒序列，并继续染色体的重排，这会进一步产生致癌的可能。当完全失去端粒时，细胞可能发生破坏或死亡。但如果危机前期的遗传混乱导致端粒酶的产生，细胞将不会完全失去它们的端粒。缩短的端粒将被保持下来。这样，遗传混乱的细胞将获得癌的永生特征。在各种人的癌症中存在端粒酶以及在正常体细胞中不存在端粒酶的现象，使人们意识到对这种酶的研究可能导致发现新的抗癌药物。因为能够阻碍端粒酶的药物可能通过使端粒缩短或消失破坏癌细胞而不影响大多数正常细胞的功能。此外，由于端粒酶存在于许多癌症中，所以端粒酶的抑制剂将对肿瘤具有广泛的疗效。

当然，上述的假设还仅仅是一种理论，因为还有一些例外的情况，如有些晚期肿瘤缺乏端粒酶活性，而最近发现有些体细胞（一些白细胞）也能产生这种端粒酶，但总的说，已有证据表明，大多数肿瘤细胞需要端粒酶以进行无限的分裂。

（吴东林）

参 考 文 献

[1] Kelman Z and O Donnell M. Ann Rev Biochem, 1995. 64:171～200
[2] Nakamis V et al. Cell, 1996. 84:137～145
[3] Herendeen D R et and Kelly T J. Cell, 1996. 84:5～9
[4] Kiefer J R et al. Nature, 1998. 391:304～307
[5] Steitz T A. Nature, 1998. 391:231～232
[6] Doublié S. Nature, 1998. 391:251～258
[7] Lohman T M and Bjornson, K P. Ann Rev Biochem, 1996. 65:169～214
[8] Marians K J. Ann Rev Biochem, 1992. 61:673～719
[9] Baker T A and Bell S P. Cell, 1998. 92:295～305
[10] Baker T A. Cell, 1995. 80:521～524
[11] Shadel G S. and Clayton D A. Ann Rev Biochem, 1997. 66:409～435
[12] Stillman B. J Biol Chem, 1994 269:7047～7050
[13] Waga S and Stillman B. Nature, 1994. 369:207～212
[14] Borowiec J A et al. Cell, 1990. 60:181～184
[15] Tsurimoto T et al. Nature, 1990. 346:534～539

[16] Tsurimoto T and Stillman B. J Biol Chem, 1991. 266:1950~1968
[17] Melendy T and Stillman B. J Biol Chem, 1991. 266:1942~1949
[18] Wold M S. Ann Rev Biochem, 1997. 66:61~92
[19] Marahrens Y and Stillman B. Science, 1992. 255:817~823
[20] Diffley J F X et al. Cell, 1994. 76:303~316
[21] Toone W M et al. Ann Rev Microbiol, 1997. 51:125~149
[22] Stillman B. Science, 1996 274:1659~1664
[23] Greider C W and Blackburn E H. Cell, 1985 43:405~413
[24] Greider C W and Blackburn E H. Nature, 1989. 337:331~337
[25] Morin G B. Cell, 1989. 59:521~529
[26] Kim N W et al. Science, 1994. 266:2011~2015
[27] Greider C W. Ann Rev Biochem, 1996. 65:337~365
[28] Lingner J et al. Science, 1997 276:561~567
[29] Meyerson M et al. Cell, 1997. 90:785~795

第五章　转　　录

转录(transcription)是指以 DNA 为模板,在依赖于 DNA 的 RNA 聚合酶的催化下,以 4 种 rNTP(ATP、CTP、GTP 和 UTP)为原料,合成 RNA 的过程。RNA 是 DNA 将遗传信息传递给蛋白质的中心环节。在有些 RNA 病毒中,RNA 也可以指导合成 RNA。

在转录过程中,DNA 双链中的一条链为模板,称为模板链,而另一条链称为编码链。转录起始于 RNA 聚合酶和启动子结合之后,转录起始的第一个碱基称为转录起始点,在 RNA 聚合酶的作用下合成 RNA,至终止子处终止。由启动子到终止子的序列称为转录单位(transcription unit)。转录起始点前面的序列称为上游(upstream),后面的序列称为下游(downstream)。起始点为 +1,上游的第一个核苷酸为 -1,其他的依次类推。

对原核生物的转录过程研究较早,了解得较详尽。近年来对真核生物的转录过程也有了较多的了解。下面我们讲述原核生物和真核生物转录过程及其特征。

第一节　转录酶和转录因子[1~13]

对 RNA 合成的酶学过程的研究始于 50 年代末,目前的资料比较充分,了解得比较详细。

一、原核生物的 RNA 聚合酶

大部分原核生物的 RNA 聚合酶的结构十分相似,但蓝藻菌和古细菌的 RNA 聚合酶比较特别。在真杆菌中,只有一种 RNA 聚合酶,负责 mRNA、tRNA 和 rRNA 的转录。细菌的 RNA 聚合酶的大小约为 9.5nm × 9.5nm × 16 nm。在聚合酶的表面,形成一条约 2.5 nm 宽的通道,或称为沟,可能是容纳 DNA 分子的场所。大肠杆菌的 RNA 聚合酶是目前了解最详细的 RNA 聚合酶。在一个大肠杆菌细胞中,大约有 7000 个 RNA 聚合酶分子,其中有 2000~5000 个酶分子在参与 RNA 的合成。在生长旺盛时,参与 RNA 合成的分子较多。RNA 聚合酶合成 RNA 的速度在 37℃为约 40 个核苷酸每秒。

1. 大肠杆菌 RNA 聚合酶

大肠杆菌 RNA 聚合酶由多亚基组成。$\alpha_2\beta\beta'\sigma$ 称为全酶(holoenzyme),分子量约为 480 kDa。σ 亚基与全酶结合疏松,应用磷酸纤维素柱很容易使之与全酶分离。解离后的部分称为核心酶(core enzyme)($\alpha_2\beta\beta'$)。全酶可以起始 RNA 的合成,之后 σ 亚基从全酶解离下来。全酶的组装过程可能如下:$\alpha_2+\beta+\beta'+\sigma$。

各亚基的功能已有了一定的了解。通过利福霉素(rifamycin)类药物和链霉溶菌素(streptolydigin)抑制转录的实验表明,β 亚基可能具有与底物(NTP 及新生 RNA 链)结合

的能力。利福霉素可阻断新生的RNA链转移至RNA聚合酶的结合位点，阻断新生RNA链的第二或第四个核苷酸的添加，从而阻断转录的起始。链霉溶菌素可抑制延伸反应。二者均是通过与β亚基的结合而发挥作用的。β′亚基可能与模板结合。肝素可与β′亚基结合而抑制转录，并且可以和β′亚基竞争DNA的结合位点。β和β′亚基提供了RNA聚合酶的活性中心，其一级结构与真核生物RNA聚合酶的大亚基有同源性。α亚基可能参与全酶的组装及全酶识别启动子。在T4噬菌体感染大肠杆菌时，α亚基的一个精氨酸残基被ADP-核糖基化，从而降低了与启动子的亲合力。另外，α亚基还参与RNA聚合酶与一些调控因子间的作用。核心酶具有与DNA非特异性结合的能力，称为松弛型结合，结合常数为10^{11}L/mol，半衰期为60分钟，这种结合是DNA的磷酸基团与蛋白质的碱性基团间的非特异性吸附作用。全酶与DNA的非特异性结合下降至结合常数为10^{7}L/mol，半衰期在1秒以下，而与启动子的结合能力大大提高，结合常数达10^{14}L/mol，半衰期长达几个小时。但全酶与不同启动子间的结合力也有百倍之差。σ亚基的功能是使全酶识别启动子并与之结合。

在大肠杆菌的T5和T7噬菌体，其RNA聚合酶只由一条多肽链组成，分子量小于100 kDa，37℃时合成RNA的速度为约200个核苷酸/秒，比大肠杆菌RNA聚合酶快得多。这表明RNA的合成可以在一条多肽链的催化下完成。而大肠杆菌的RNA聚合酶却有多个亚基组成。这种结构的复杂性决定了其功能的复杂性。T4和T7噬菌体的RNA聚合酶只能识别噬菌体DNA的少数几个启动子，而大肠杆菌RNA聚合酶却能识别1000个以上转录单位的启动子。有些转录单位可被直接转录；有些需要蛋白质辅助因子的参与。部分辅助因子只参与一种转录单位的转录；另一部分则参与多种转录单位的转录。有时，噬菌体还可改变宿主RNA聚合酶的亲合力，使其不能识别宿主DNA的启动子，而识别噬菌体DNA的启动子，从而转录噬菌体的基因。由此可见，大肠杆菌RNA聚合酶多亚基的复杂结构主要是为酶提供了与多种因子相互作用的能力，从而可识别多种启动子。

在转录过程中，RNA聚合酶需要与其他蛋白质辅助因子共同作用，才能保证转录的顺利进行。如转录终止时的终止因子和抗终止因子。σ亚基实际也可被看作一种辅助因子，只是习惯上将其看成RNA聚合酶的一个亚基，因此又可称为σ因子(σ factor)。延伸过程中同样需要蛋白质辅助因子的参与。

2. σ因子

上已提及，σ因子在RNA聚合酶识别启动子的过程中起关键作用。在大肠杆菌中，RNA聚合酶要负责所有基因的转录，也就是说要能识别所有转录单位的启动子。并且，在不同的生长时期，基因的表达是不同的。因此，RNA聚合酶在识别启动子时必须有很强的灵活性。这种灵活性不可能体现在核心酶上，而只能由σ亚基来承担。

在很长一段时间内，人们一直认为大肠杆菌只有一种σ因子(σ^{70})。目前已发现了至少4种σ因子(表5-1)σ因子是根据其分子量命名的。σ^{32}(σ^{H})和σ^{54}(σ^{N})因环境变化而激活。在环境温度升高时，σ^{32}使热激(heat shock)基因表达；在氮缺乏时，σ^{54}使利用其他氮源相关的基因表达。σ^{28}(σ^{F})负责基因的表达，但表达的水平受环境因素的影响。

表 5-1 大肠杆菌的 σ 因子及其特性

基因	分子量	用途	-35 区	识别序列间隔	-10 区
rpoD	70 kDa	普通	TTGATC	16~18 bp	TATAAT
rpoH	32 kDa	热激	CCCTTGAA	13~15 bp	CCCGATNT
rpoE	24 kDa	热激	不详	不详	不详
rpoN	54 kDa	氮缺乏	CTGGNA	6 bp	TTGCA
fliA	28 kDa	芽孢	CTAAA	15 bp	GCCGATAA

枯草杆菌(*B. subtilis*)的 RNA 聚合酶具有与大肠杆菌相同的结构($\alpha_2\beta\beta'\sigma$),它有 10 种左右的 σ 因子。有些 σ 因子存在于生长期细胞,有些只出现在噬菌体感染或由正常生长状态转变为芽孢状态等特殊情况下。σ^{ABCDHL}负责生长期基因的转录,σ^{EFGK}负责孢子形成期基因的转录。正常生长状态下,其 σ 因子的分子量为 43 kDa,写做 σ^{43}或 σ^{A}。它所识别的启动子序列与大肠杆菌的 σ^{70}相同。其他的 σ 因子在细胞中的含量较少,且识别的启动子序列也有所不同。随着新的基因序列的发现,可能会发现新的 σ 因子。

σ 因子是通过识别启动子上的某一序列来控制 RNA 聚合酶与启动子的结合的。在启动子的结构中,有两处保守序列,位于 -35 区和 -10 区。在每一组启动子中,或是两处保守序列均有所差异,或是一处保守序列有所不同。而这些不同的保守序列能被不同的 σ 因子所识别。σ 因子可以直接与启动子 -35 区至 -10 区的 DNA 序列相结合,但还不清楚一条小肽如何结合 20 多个 b 导致其失去与 RNA 聚合酶结合的能力,但 σ 因子的突变可恢复 RNA 聚合酶与突变启动子的结合。经类似的研究确定,大肠杆菌的 σ^{70}有 4 个保守区域,其中 2 区和 4 区与 -35 区和 -10 区的保守序列结合;2 区的一部分能与单链 DNA 结合,可能与解链有关;361 至 390 氨基酸残基负责与核心酶的结合;N 端区域去除后的 σ 因子能特异的与启动子结合,因而其功能是阻止游离的 σ 因子与启动力子的结合。其他序列的功能还不十分清楚。σ 因子的二级结构属于 α 螺旋。

α 螺旋识别 DNA 的特异序列是一种常见的方式。在 σ^{70}和 σ^{A},相隔 3、4 个氨基酸残基远的氨基酸残基位于螺旋的同一面,与相邻的碱基相互作用。如苏氨酸和精氨酸,分别与 -10 区保守序列的头两个碱基 A(-13 位)和 T(-12 位)相互作用。

σ^{54}是一个比较特殊的因子。它所识别的启动子序列和其他的 σ 因子有明显的区别,一个保守序列位于 -10 区,另一个位于 -20 区。在没有核心酶的情况下,σ^{54}因子就能与启动子结合。

3. σ 因子的更替对转录起始的调控

在大肠杆菌中,RNA 聚合酶转录了所有的基因,其 σ 因子可识别多种启动子。对特异启动子的识别需要特异辅助因子的参与。在细菌受到外界环境的急剧影响或芽孢菌生活方式改变时,会产生 σ 因子更替的现象。

在环境温度升高时,大肠杆菌会产生热激状态,所转录的基因发生变化。在多种生物,包括真核生物和原核生物均有这种现象的发生。当温度升高时,正常状态下表达的基因会关闭或表达水平下降,而热激基因(heat shock gene)表达。这些蛋白能提高菌体对高

温的抵抗能力，其中有些蛋白质属于分子伴侣(chaperons)。在大肠杆菌中，*rpoH* 的表达促发其他热激基因的表达，其产物为 σ^{32}，能识别热激基因的启动子。当温度升高时，σ^{32}在细胞内的浓度升高，温度恢复正常时，σ^{32}的浓度下降。温度升高时，细胞内会有一些蛋白质发生变性，从而为 σ^{32}的表达提供了信号。调控热激蛋白表达的另一种因子为 σ^{E}，由细胞周间质或膜的变性蛋白提供信号。引起 σ^{E}表达的温度变化比引起 σ^{32}表达的变化更加剧烈。

氮的缺乏也会引起大肠杆菌 σ 因子的更替。大肠杆菌细胞中含有少量的 σ^{54}因子，当铵缺乏时，该因子所识别的基因表达，可利用其他的氮源。大多种细菌在进化过程中均保留了这种机制。σ^{E}也属于这类因子，它可引起与化学趋向性和鞭毛结构有关基因的表达。枯草杆菌中的 σ^{D}与大肠菌中的 σ^{E}功能相同，与鞭毛和运动有关的基因表达有关。在很多细菌中，均有这类 σ 因子的存在。

在枯草杆菌，σ 因子的更替控制着生活方式的改变。枯草杆菌共有约 10 种 σ 因子。其中有些出现于正常生活期，有些出现于噬菌体感染或是芽孢中。枯草杆菌的 RNA 聚合酶具有与大肠杆菌相同的结构，正常生活期的 RNA 聚合酶的 σ 因子为 $\sigma^{43}\sigma^{A}$，识别与 σ^{70}相同的启动子保守序列。还有其他种类的 σ 因子，识别不同的启动子，但含量很少。

在噬菌体感染时，细菌会改变所表达的基因是一种普遍的现象。这是因为噬菌体 RNA 聚合酶的表达或是噬菌体编码的辅助因子对细菌 RNA 聚合酶的调控作用。下面介绍一下噬菌体 SPO1 感染时细菌基因表达的变化。

SPO1 感染后基因表达的改变分为三个阶段，首先是早期基因的表达，4～5 分钟后中期基因表达，8～12 分钟后晚期基因表达。早期基因具有与宿主基因相同的启动子，可由宿主的聚合酶($\alpha_2\beta\beta'\sigma^{43}$)转录。中期和晚期基因的表达受三个调控基因的影响，分别为 *28*、*33* 和 *34*。早期基因 *28* 编码一种 σ 因子(gp28)，与核心酶组成全酶转录中期基因；中期基因 *33* 和 *34* 的表达产物(gp33 和 gp34)代替 gp28，表达晚期基因，形成了一种瀑布调控方式。gp28 基因的突变使中期基因不能表达，它的分子量为 26 kDa，替代宿主的 σ 因子后，使得 RNA 聚合酶不能转录宿主的基因，而转录中期基因。目前还不清楚 gp28 是如何代替 σ^{43}的。基因 *33* 和 *34* 的突变可阻止晚期基因的表达，其基因产物分别为 13 kDa 和 24 kDa，可代替 gp28，起始晚期基因的表达。

在生活周期的转换过程也是通过 σ 因子的更替完成的。在对数生长期以后，由于营养物质的缺乏，细菌转向芽孢生长期。DNA 完成复制，基因组位于细胞的一端，最后细胞由坚实的孢衣包围，由原来的一个细菌形成一个芽孢菌和一个正常细菌。这一过程要经历约 8 小时。由正常细菌转变为芽孢菌，菌体的代谢发生了急剧的变化，有些基因关闭，有些基因开启表达，但大部分基因持续表达。在芽孢菌中，约有 40% 的特异 mRNA。芽孢菌的 RNA 聚合酶有特异的 σ 因子。

SpoOA 的磷酸化是芽孢形成过程的开端。磷酸基团经过几次传递最终使 SpoOA 磷酸化。这一过程需要几种基因的产物。SpoOA 是一个转录调控因子，其活性受磷酸化状态的控制。在磷酸化时，可使细菌的 RNA 聚合酶(σ 因子为 σ^{43})转录 σ^{F}的基因。

二、真核生物的 RNA 聚合酶

在真核生物细胞中，共有三类 RNA 聚合酶，即 RNA 聚合酶 I、II 和 III，分别负责三类基因的转录(表 5-2)。分类的依据最早是根据它们从 DEAE 纤维素柱上洗脱下来的先后顺序。但后来发现不同真核生物的三种 RNA 聚合酶的洗脱顺序并不相同，于是采用了根据对鹅膏蕈碱(α-amanitine)的敏感性不同来分类的方法。它们在细胞核中的定位不同，对鹅膏蕈碱的敏感性有差异。动物、植物和昆虫的 RNA 聚合酶 I 最不敏感，RNA 聚合酶 II 最敏感。不同来源的 RNA 聚合酶 III 有些差异，动物细胞的 RNA 聚合酶 III 的敏感性介于 I 和 II 之间，受高浓度的鹅膏蕈碱的抑制；酵母和昆虫的 RNA 聚合酶 III 不受抑制。RNA 聚合酶的抑制剂在分类上有一定的意义。

表 5-2 真核生物的 RNA 聚合酶

酶	定位	转录产物	相对活性	对鹅膏蕈碱的敏感性
RNA PolI	核仁	rRNA	50%～70%	不敏感
RNA PolII	核质	核 RNA	20%～40%	敏感
RNA PolIII	核质	tRNA	10%	不同种类的敏感性不同

在细胞核中，RNA 聚合酶 I 位于核仁，活性所占比例最大，负责 rRNA(5.8 S、18 S 和 28 S)的转录。由于 rRNA 占总 RNA 的比例最大，所以 RNA 聚合酶 I 负责了大部分细胞 RNA 的转录。RNA 聚合酶 II 位于核质，活性所占比例次于 RNA 聚合酶 I，主要负责不均一核 RNA(heterogeneous nuclear RNA，hnRNA)和 mRMA 前体的转录。RNA 聚合酶 III 也位于核质，活性所占比例最小，负责 tRNA、5 SRNA、*Alu* 序列和其他小 RNA 的转录。

真核生物的 RNA 聚合酶的分子量均很大，为 500 kDa 或更大。由 8 至 14 个亚基组成。经纯化的 RNA 聚合酶具有以 DNA 为模板合成 RNA 的能力，但不能正确的选择启动子。目前尚不能完成 RNA 聚合酶的体外重建，所以无法确定哪些亚基是活性所必需的，哪些是活性所非必需，是不是所有的亚基都是活性所必需的。目前了解最为详细的是酿酒酵母，所有的亚基均已找到了编码的基因，其中 10～11 个亚基是活性所必需的。

酿酒酵母的 RNA 聚合酶 II 有 10 多个亚基组成，其中最大的三个亚基与细菌的 RNA 聚合酶有同源性，并可能由此三个亚基组成基本的催化活性。其中的两个亚基可能具有催化位点。另有三个亚基是 RNA 聚合酶的共同亚基，在 RNA 聚合酶 II 和 III 中均存在。多数亚基是单拷贝 RNA 聚合酶 II 的最大的亚基，有一个 C 末端结构域(carboxy-terminal domain，CTD)，含一保守序列的多个重复，保守序列为 Tyr-Ser-Pro-Thr-Ser-Pro-Ser，称为 C 端重复七肽，只出现于 RNA 聚合酶 II 中。在酵母中有 26 个这样的重复，在哺乳动物有 50 个。重复的数目对于酶的活性十分重要，缺失一半以上的重复属致死变

异。CTD 中的 Ser 和 Thr 可被高度磷酸化，磷酸化的 RNA 聚合酶称为 II_a，非磷酸化的称为 II_o。CTD 参与了转录的起始。

线粒体和叶绿体的 RNA 聚合酶活性较小，与细菌的 RNA 聚合酶更为相似。其功能也较真核生物的 RNA 聚合酶简单，只需转录为数不多的几个特定的基因，而不需要更为复杂的功能来适应环境的变化。

第二节　启 动 子[5,7,14~17]

启动子(promoter)是指 DNA 分子上被 RNA 聚合酶识别并结合形成起始转录复合物的区域，它还包括一些调节蛋白因子的结合位点。无论是原核生物还是真核生物，启动子是控制转录起始的序列，并决定着某一基因的表达强度。与 RNA 聚合酶亲和力高的启动子，其起始频率和效率均高。启动子的 DNA 序列本身就可以提供特定的信号，而转录区域的 DNA 序列要转变成 RNA 或蛋白质后才能体现出它所贮存的信息。原核生物和真核生物的启动子的结构还有些差异。

一、原核生物的启动子

应用 RNA 聚合酶来确定可与之结合的 DNA 序列，可用于对启动子的研究。在启动子中，应该含有可与 RNA 聚合酶结合的序列。在原核生物的基因组中，能提供一定信息的最小 DNA 片段为 12 bp，即能被 RNA 聚合酶识别的最小长度为 12 bp。这 12 bp 的长度不一定是连续排列的，可以被其他的序列隔开，而且，在这种情况下，这些间隔序列的长度本身就可以提供部分信息。因此，被隔离开的序列长度之和可以小于 12 bp，就能被 RNA 聚合酶识别。

研究启动子特性的方法之一是比较不同启动子的序列，从中发现一些同源性比较强的部分，又称保守序列。同源性强的序列应该是启动子的功能所必需的。通过研究发现，保守序列并不是在任何的启动子中均是相同的，存在着一定的灵活性，但一个启动子与保守序列的差别不应超过 1 到 2 个碱基。在表示保守序列时，同时标明该位置保守碱基的出现频率。

现已比较了 100 多个大肠杆菌的启动子，在与 RNA 聚合酶相互作用的序列中，几乎没有发现 60 bp 以上的比较保守的序列。多数结合位点的序列没有同源性。但发现了一些短的保守序列，并且是启动子的功能所必需的。这种短的保守序列也是真核生物和原核生物基因组中调控位点的典型特征。

1. 原核生物启动子的结构

在原核生物的启动子中，有 4 个区域：转录起始点、-10 区、-35 区、-10 与 -35 之间的序列。

(1)转录起始点　转录起始点在多数情况(>90%)下为嘌呤，常见的序列为 CAT，A 为转录起始点。但 CAT 的保守程度还不够高，因而不能做为固定的转录起始点序列。

(2)－10 区　　在转录起始点的上游紧接着一 6 bp 的保守序列，在几乎目前已知所有的启动子中均存在。保守序列的中心位于转录起始点上游约－10 bp 处，在不同启动子中有所差别，在－18 bp 到－9 bp 之间。因此这一保守序列又称为－10 序列。其一致序列为 TATAAT，由于每个碱基的出现频率不同，又可写为 $T_{80}A_{95}T_{45}A_{60}a_{50}T_{96}$。数字表示该位置碱基在目前所有的启动子中出现的频率(%)。大写字母表示频率大于 54%。小写字母表示该位置的碱基保守性差些，但比随机分配的概率要高。如果该位置的碱基没有保守性，则用 N 表示。前两位的 AT 和最后一位的 T 的保守性最强，预示着这几个碱基在与 RNA 聚合酶的作用上可能有十分重要的作用。

由于该保守序列最早由 Pribnow 提出，所以又称为 Pribnow 框。它是 RNA 聚合酶的牢固结合位点，又称结合位点，或称为解链区。该区域富含 AT 对，熔点较低，在 RNA 聚合酶的诱导下易于首先熔解，使封闭复合物转变为开放复合物，便于转录的起始。

(3)－35 区　　在转录起始点上游　35 bp 处，有另一个保守序列，称为－35 序列。其保守序列为 TTGACA，由于每个碱基的出现频率不同，又可写为 $T_{82}T_{84}G_{78}A_{65}C_{54}a_{45}$。该保守序列又称 Sextama 框。RNA 聚合酶的 σ 因子可以识别该位点，所以称作识别位点。RNA 聚合酶先识别这一区域并与之结合，然后再与结合位点相互作用。该区域是启动子强弱的决定因素。

(4)－10 区和－35 区间的距离　　－35 区和－10 区之间的距离在 90% 的原核生物启动子为 16 到 18 bp。少数情况下，该距离可少至 15 bp，多至 20 bp。该区域内的碱基序列并不重要，但碱基的数目即该距离的长短却是至关重要的。适宜的距离可以为 RNA 聚合酶提供合适的空间结构，便于转录的起始。

分析大量启动子的结构得出，典型的原核生物的启动子的结构为－35 区、16～19 bp 的间隔区、－10 区和转录起始点，－10 区到转录起始点的距离为 7 bp。

并不是所有的启动子都具有典型的结构。有些启动子缺少其中的某一结构。有时只有 RNA 聚合酶本身并不足以与启动子结合，还需要其他的辅助因子。

另外，转录起始点左右的碱基也可影响转录的起始。＋1 至＋30 的转录区可影响 RNA 聚合酶对启动子的清除，从而影响启动子的强度。

2. 启动子功能的研究方法

研究启动子功能的主要方法是启动子的突变。使启动子的碱基序列发生变化，可以改变启动子的强弱，也就是说增强或减弱该启动子所控制的结构基因的转录。大多数突变是使启动子的功能减弱或消失，这种突变称为下降突变；少数突变能使启动子的功能增强，称为上升突变。应当注意，下降或上升是突变前后同一启动子的变化情况，在不同的突变体或同一突变体不同的启动子中，情况是不同的。发生在一个菌株的某一启动子的突变，不一定发生在另一菌株。

目前所知，几乎能影响启动子功能的所有点突变均发生在两个保守区域内(－10 区和－35 区)，极少数的发生在两个保守区域上游的几个碱基内，其他位置的点突变几乎不影响启动子的功能。此外，在两个保守区域间的插入或缺失突变改变了二者的距离，从而改变启动子的强弱。有趣的是，使两个保守序列更加趋于保守或使两保守区域间的距离接近于 17 bp 的突变多为上升突变，使保守区域的保守性下降或两保守区间的距离偏离

17 bp 的突变多为下降突变。下降突变多发生于保守性强的碱基。但也有些例外的情况并不遵守这一规律，说明保守区内碱基的功能并不十分确定，也表明我们并不能十分准确的从启动子的结构来预测其功能的强弱。实际上每个启动子的结构都有差异，即使某一碱基在某一位置是保守的，它的相邻碱基也有差异。在另一种情况下，某一启动子某一位置的碱基尽管不是保守碱基，但对这一启动子的功能仍有重要的作用。如将这一碱基换到另一个启动子中，则会使该启动子失去功能。在有些启动子中，某些保守位置上并不含保守的碱基，但它们仍有功能。

要想准确测定启动子的突变效应，必须在体外测定 RNA 聚合酶与野生型和突变型启动子的亲合力。不同启动子与 RNA 聚合酶的亲合力有很大的差别，最多可相差 100 倍。这也可说明不同的基因在细菌体内表达的强度有很大差异。通过上述分析，我们可以弄清楚突变的机制，看其是改变启动子与 RNA 聚合酶的亲和力，还是酶可以与启动子结合但不能起始转录，或是某些辅助因子发生了变化。

通过分析动力学常数可知，－35 区的下降突变降低了封闭复合物的形成率，但不影响向开放复合物的转变。－10 区的下降突变不影响封闭复合物的形成，但减缓转变为开放复合物的速度。该结果表明，－35 区的功能是提供 RNA 聚合酶的识别信号，－10 区是使封闭复合物转变为开放复合物。

在有些启动子中，并不含 －10 区和 －35 区这样的保守序列，因而不能被 RNA 聚合酶所识别，必须在其他的蛋白辅助因子的共同作用下，才能完成转录的起始。有关这方面的情况，目前了解的还不十分详细。

3. RNA 聚合酶与启动子的结合

研究 RNA 聚合酶与启动子间的相互结合，是揭示转录起始机制的重要手段。研究 RNA 聚合酶与 DNA 间的识别及结合常采用足迹法(footprinting)。足迹法的基本原理是应用内切酶部分水解 DNA。控制反应条件，可使 DNA 在每个碱基均有机会被切断。先将大分子 DNA 一条链的末端进行放射性标记，经内切酶部分酶切，应用凝胶电泳分离，再进行放射自显影。小的片段移动较快，走在凝胶的前面，大的片段移动较慢，走在后面。这和 DNA 序列测定的原理相同。如果先使 DNA 与 RNA 聚合酶结合，再用内切酶水解，与 RNA 聚合酶结合的 DNA 会受到保护而免受水解。同时将未结合 RNA 聚合酶的 DNA 作为对照，面后将水解产物进行凝胶电泳和放射自显影。在自显影图上，对照组会出现连续的条带，而实验组在与 RNA 聚合酶结合的区域会无条带出现。和对照组相比较，可直接阅读出这一区域的 DNA 序列，也就是 DNA 和 RNA 聚合酶结合区域的序列。经测定，RNA 聚合酶与启动子结合的区域为 ＋20 至 －50，这段序列包括了结合及起始所需的序列。在实验中发现，DNA 的两条链受 RNA 聚合酶的保护程度不同，特别是在结合位点的边缘。说明 RNA 聚合酶与两条链的结合是不对称的，也表明转录只需一条模板链。

另一种研究方法是采用修饰碱基的方法。可先用碱基修饰试剂对 DNA 和 RNA 聚合酶的复合物进行修饰，未与 RNA 聚合酶结合的 DNA，其相应的碱基被修饰，而结合部分的碱基则不被修饰。再通过上述方法来分析未被修饰的碱基，即为与 RNA 聚合酶结合的部位。或是先将 DNA 修饰，然后再与 RNA 聚合酶结合，分析那些不能与 RNA 聚合

酶结合的片段。这些分析方法可以确定与 RNA 聚合酶紧密结合的位点。

经上述分析，大多数与 RNA 聚合酶紧密结合的位点集中在启动子的 -35 区至 -10 区之间。在这一区域，碱基的修饰可以阻止与 RNA 聚合酶的结合。与经突变研究所确定的功能区域相比，RNA 聚合酶的紧密结合位点区域要长 12～15 bp，但仍集中在同一区域。如图 5-1 所示。

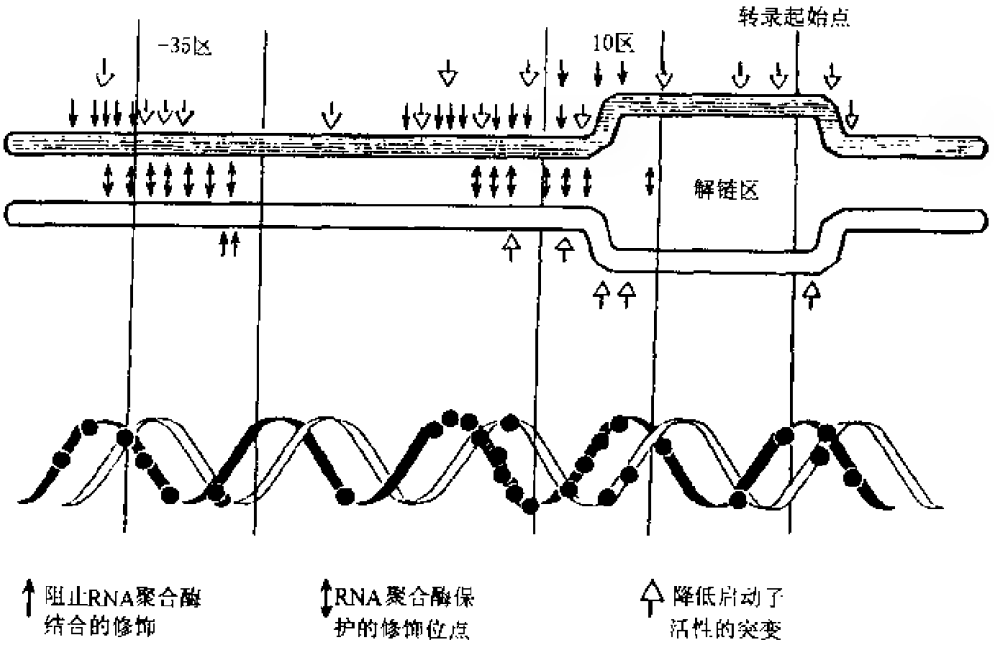


图 5-1　与 RNA 聚合酶紧密接触的位点

值得注意的是，在不同的启动子中，有些紧密接触位点是位于同一个位置上，而该位置的碱基却不同。这表明 RNA 聚合酶与 DNA 的结合存在一个共同的机制，而且对结合位点的碱基种类无特殊要求。这也解释了为什么结合位点和引起启动子功能变化的突变位点并不统一。

应用修饰碱基的方法还可确定启动子中处于熔解状态的部分。有些修饰试剂不能修饰处于双链状态的碱基，但可修饰处于单链状态的碱基。这样，用使 A 或 C 甲基化的试剂很容易鉴别出 -9 至 +3 处于单链状态。

十分有趣的是，在 -10 区上游的所有紧密结合位点均位于一条 DNA 链上，在双链 DNA 中，这些紧密结合位点处于 DNA 的一侧。这可使 RNA 聚合酶从 DNA 的一个侧面接近并识别启动子并与之结合。另外，大部分的紧密结合位点位于编码链上，少数位于模板链上，且都处于 -10 区的下游。可能是 DNA 双链熔解后，使 RNA 聚合酶得以识别模板链上的位点并与其结合。

转录的起始需要一段 DNA 的解链也可以解释另外一种现象，即 RNA 聚合酶对处于超螺旋的 DNA 起始效率高，因为处于超螺旋的 DNA 解链所需的自由能少。因此，在原核生物中，可以通过控制 DNA 的超螺旋状态来调解启动子的活性，如 DNA 旋转酶(DNA gyrase)能增加 DNA 的负超螺旋，拓扑异构酶 I(topoisomerase I)能消除负超螺旋。抑制或

经突变改变这两种酶的活性可影响转录。其机制尚未完全清楚，细菌中存在一种机制来使DNA的超螺旋程度保持在一定的范围内。但目前可以肯定，有些启动子的活性受超螺旋状态的影响，多数情况下负超螺旋有助于转录的起始，但也有些启动子并不受超螺旋的影响。造成这种差异的原因可能是有些启动子的序列易于熔解，对超螺旋的依赖性差些，而有些启动子序列熔解难一些，必须借助超螺旋来降低自由能。也有些启动子的活性随超螺旋程度的增加而减小，目前尚不能解释其原因。在转录的过程中，也存在拓扑学的问题。

二、真核生物的启动子

真核生物细胞中有三种转录方式，分别由三种RNA聚合酶(I、II和III)催化，因此有三种启动子。根据启动子的不同，将真核生物的基因分为三类，即I类、II类和III类基因。这三种基因分别由三种启动子控制，它们在结构上各有特点。

1. RNA聚合酶I的启动子的结构

和其他两类启动子相比，RNA聚合酶I的启动子间的差异最小。因为RNA聚合酶I只转录rRNA一种基因，包括5.8 S、18 S和28 S rRNA。三种rRNA的基因(rDNA)成簇存在，共同转录在一个转录产物上，然后经加工成为三种rRNA。

RNA聚合酶I的启动子主要由两部分组成。目前了解较清楚的是人的RNA聚合酶I的启动子。在转录起始位点的上游有两部分序列。核心启动子(core promoter)位于-45至+20的区域内，这段序列就足以使转录起始。在其上游有一序列，从-180至-107，称为上游调控元件(upstream control element，UCE)，可以大大的提高核心启动子的转录起始效率。两个区域内的碱基组成和一般的启动子结构有所差异，均富含G.C对，两者有85%的同源性。

2. RNA聚合酶II的启动子的结构

RNA聚合酶II主要负责蛋白质基因和部分核小RNA(small nuclear RNA，snRNA)的转录，其启动子结构最为复杂。RNA聚合酶II单独并不能起始转录，必须和其他的辅助因子共同作用形成转录起始复合物后才能起始转录。RNA聚合酶II的启动子位于转录起始点的上游，由多个短序列元件组成。该类启动子属子通用型启动子，即在各种组织中均可被RNA聚合酶II所识别，没有组织特异性。通过对β珠蛋白基因转录起始点上游的100个碱基进行替代试验发现，在HeLa细胞中，大多数的突变并不影响启动子起始转录的能力。其中有三个下降突变，发生在三个短序列元件上，远离转录起始点的短序列元件突变对转录起始的影响比接近起始点的短序列元件要大。只发现了一例上升突变。

经过比较多种启动子，发现RNA聚合酶II的启动子有一些共同的特点，在转录起始点的上游有三个保守序列，又称为元件(elememt)。

(1)帽子位点(cap site)　帽子位点又称转录起始位点，其碱基大多为A(指非模板链)，这与原核生物相似。

(2)TATA框(TATA box)　TATA框位于-30处，又称Hogness框或Golderg-

Hogness 框。一致序列为 TATAA(T)AA(T)经突变试验分析，它是三个元件中转录起始效率最低的一个。虽然有些 TATA 框的突变不影响转录的起始，但可改变转录起始位点。这说明 TATA 框具有定位转录起始点的功能。将 TATA 框反向排列，也可降低转录的效率。TATA 框周围为富含 GC 对的序列，可能对启动子的功能有重要影响。它和原核生物的启动子有些相似。TATA 框具有选择起始点的功能。在有些启动子中缺少 TATA 框。

(3)CAAT 框(CAAT box)　CAAT 框位于转录起始点上游的 −75 bp 处，一致序列为 GGC(T)CAATCT，因其保守序列为 CAAT 而得名。虽名为 CAAT 框，头两个 G 的作用却十分重要。它是最先被人们发现的转录起始元件。虽名为 CAAT 框一般位于 −80 bp 左右，但离转录起始位点距离的长短对其作用影响不大，并且正反方向排列均能起作用。CAAT 框内的突变对转录起始的影响很大，说明它决定了启动子起始转录的效率及频率。对子启动子的特异性，CAAT 框并无直接的作用，但它的存在可增强启动子的强度。

(4)GC 框(GC box)　GC 框位于 −90 bp 附近，核心序列为 GGGCGG，一个启动子中可以有多个拷贝，并且可以正反两个方向排列。GC 框也是启动子中相对常见的成分。

另外，在有些启动子中还发现了其他的元件，如八聚核苷酸元件(octamer element，OCT element)，一致序列为 ATTTGCAT；κB 元件，一致序列为 GGGACTTTCC；ATF 元件，一致序列为 GTGACGT。在转录起始点下游也有一些与启动子功能有关的元件。各元件间的距离对启动子的功能没有太大影响，不同启动子中各元件的距离差异很大。但如果距离太近(小于 10 bp)或太远(大于 30 bp)，就会影响启动子的功能。

上面讲述了组成启动子的基本元件。在不同的启动子中，这些元件的组合情况是不同的。在这些启动子中，共有 4 种元件，分别为八聚核苷酸元件、CAAT 框、GC 框和 TATA 框。这些元件在三种启动子中的数目、位置和排列方向均有差异，没有一个元件在三个启动子中都存在。在 SV40 的早期启动子中，含有 6 个 GC 框；在胸腺嘧啶激酶的启动子中，有 1 个八聚体元件、2 个 GC 框、1 个 CAAT 框和 1 个 TATA 框；在 H2B 的启动子中，含有 2 个八聚体元件、2 个 CAAT 框及 1 个 TATA 框。GC 框和 CAAT 框在启动子中的排列正反方向均有，但启动子仍然能只在一个方向(下游方向)上起始转录。在启动子中，每种元件与转录起始点的距离不同。将 β-珠蛋白和胸腺嘧啶激酶启动子的相应元件互相交换，形成的杂交启动子的功能没有变化。各种元件将相应的蛋白因子结合到启动子上，而这些蛋白因子共同组合成起始复合物。蛋白因子间的相互作用决定了转录的起始。

上已提及，在有些启动子的结构中并不含有 TATA 框。在这种启动子中，也有一些启动子功能必需的元件。在转录起始点周围，尽管碱基的保守性较差，但也可发现 CA 的出现概率高些，CA 被几个嘧啶碱基包围。经计算机分析上百个启动子的序列，发现 +3 位的 T 比较保守。经过对 80 个不含 TATA 框的启动子进行突变实验，得到一个比较保守的序列 PyPyA + 1nt/APyPy。进一步研究证实，这一保守序列是启动子必需的，可能在不同组织中被一种蛋白因子所识别。在上述保守序列中，+1 位的 A、+3 位的 A 或 T 和 −1 位的嘧啶碱基是十分关键的，但只有 CANT 是不够的，还必须有几个嘧啶碱基。在保守序列中，不要求 4 个嘧啶碱基均存在启动子才有活性，只是 4 个嘧啶碱基都存在时能增

强启动子的活性，令人不解的是，这样一个序列如何保证其功能的发挥。在基因组中，会有很多类似的结构。至于哪些能作为起始信号，可能与它们所处的位置有关，即只有处于启动子特定的上下文序列中才能发挥作用。

这里有必要介绍一下起始子(initiator, Inr)。起始子首次由 Grosschdl 和 Birnstiel 提出，用于描述海胆组蛋白 H2A 基因中一包括转录起始点在内的 60 bp 长的 DNA 片段，缺失这一片段时，启动子能正常起始转录，但强度下降了 4 倍。在含或不含 TATA 框的启动子中，均有起始子发现。上面所讲述的，实际上是起始子的一些特点。

在不含 TATA 框的启动子中，转录起始位点的下游也有一些元件，对转录的起始非常重要。实验证明在 AdML 的启动子中，+7 至 +33 这段片段是启动子活性所必需的，TFIID 可保护其免受 DNaseI 的降解，说明可与之结合。另外，还发现帽子结合蛋白(cap binding protein, CBP)可与 +10 左右的区域结合；TFII－I 和 USF 可与 +45 左右的区域结合。在鼠的 *TdT* 启动子中，缺失 +33 至 +58 可使启动子完全失去活性。在果蝇基因的不含 TATA 框启动子的下游，发现了一保守序列 A/GGA/TCGTG，称为下游启动子元件(downstream promoter element, DPE)，可与高度纯化 TFIID 结合。并且，DPE 对于含 Inr 但不含 TATA 框的启动子的活性十分重要，但对含 TATA 框的启动子却无作用。

现已发现，RNA 聚合酶 II 的启动子有的只有 TATA 框；有的只有 Inr；有的二者均有；有的二者均没有。各种启动子具体的结构与功能还有待于进一步的研究。

3. RNA 聚合酶 III 的启动子的结构

RNA 聚合酶 III 转录 5 S rRNA、tRNA 和部分 snRNA 的基因。这三种基因的启动子的结构是不同的。因此，RNA 聚合酶 III 要想识别不同的启动子，必须和其他的辅助因子共同作用。RNA 聚合酶 III 的启动子可分为两类，识别方式不同。5 S rRNA 和 tRNA 基因的启动子位于转录起始点的下游，称为内部启动子(internal promoter)。snRNA 基因的启动子位于转录起始点的上游，和其他基因的启动子比较相似。启动子含有可被辅助因子识别的特殊序列，只有辅助因子与相应的序列结合后，RNA 聚合酶才能与启动子结合，从而起始转录。

内部启动子最先发现于非洲爪蟾的 5 S rRNA 基因。开始人们试图在转录起始点的上游寻找启动子，但都没能实现。在缺失试验中，即使将转录起始点的所有上游序列都缺失，转录仍能起始。因此便开始转录起始点下游序列的缺失试验，在 +55 bp 以前，转录仍能起始，只不过合成的 RNA 比正常的 5 S rRNA 短些。当缺失至 +55 bp 以后，转录便不能起始。用相同的方法从 5 S rRNA 基因的另一端开始进行缺失，当缺失进行至 +80 bp 时，转录不能起始。这说明 +55 bp 至 +80 bp 这段序列与转录起始有关，即 5 S rRNA 基因的启动子位于基因内部的 +55 bp 至 +80 bp，该启动子可使 RNA 聚合酶 III 在其上游的 55 bp 处起始转录。野生型的转录起始位点是十分固定的，当该起始点缺失时，会选择离 55 bp 最近的嘌呤碱基作为转录起始点。

进一步研究确定，内部启动子可分为两类，每类启动子均含有两个短序列元件。两个短序列元件间由其他序列隔开。第一类内部启动子含 A 框(box A)和 C 框(box C)；第二类内部启动子含 A 框和 B 框(box B)。两个保守区域间由其他序列隔开。第二类内部启动子的 A 框与 B 框间的间隔序列长短差异很大，但如果间隔序列过短，会影响启动子的

功能。转录起始点也可影响转录起始的效率，紧接转录起始点的上游序列的突变会影响转录的起始。因此可以推断，内部启动子起被 RNA 聚合酶 III 识别的作用，转录起始点附近的序列控制转录起始的效率。原核生物的 tRNA 基因的启动子也由 A 框和 B 框组成，但位于转录起始点的上游，不属于内部启动子。

RNA 聚合酶 III 的第三类启动子位于转录起始点的上游，含三个短序列元件，分别为 TATA 框、近端序列元件(proximal sequence element，PSE)和八聚核苷酸元件。有一部分 snRNA 的基因由 RNA 聚合酶 II 转录，它们的启动子也有相似的结构，三个元件也有相似的功能。TATA 框决定着启动子与两种 RNA 聚合酶的作用。RNA 聚合酶只需要转录起始点加上一段含 TATA 框的短序列就能起始转录，但 PSE 和 OCT 元件能大大提高其转录效率。PSE 对 RNA 聚合酶 II 的启动子来说是必需的，对 RNA 聚合酶 III 的启动子只起刺激作用。

RNA 聚合酶 III 的启动子结构如图 5－2 所示。

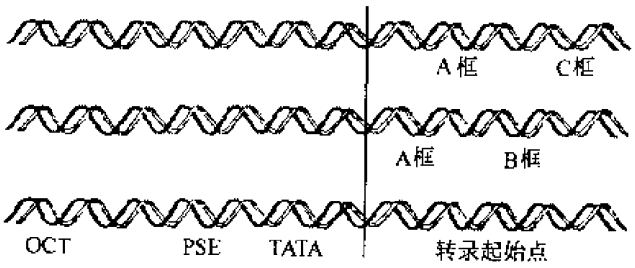


图 5－2　RNA 聚合酶 III 的启动子的结构

4．研究真核生物启动子结构与功能的方法

研究启动子结构与功能的方法主要有缺失、点突变和足迹法。在分析得到了启动子的功能序列后，还要弄清与之结合的蛋白质及两者间的相互作用。在研究真核生物启动子的结构与功能时，常采用下列方法。

(1)*卵细胞系统*(oocyte system)　该方法是将 DNA 直接注射入爪蟾(*X. laevis*)卵细胞的细胞核，分析和观察 RNA 的转录情况。该方法的局限性在于试验条件受卵细胞内条件的限制。可以用来分析 DNA 片段的特性，不能用于分析蛋白质因子与 DNA 间的结合。

(2)*转染系统*(transfection system)　将外源 DNA 导入转染的细胞并使之表达。表达可分为瞬时表达(transient expression)和整合表达(integrant expression)。由于转录是在细胞内完成的，可以看成是一种体内试验系统。但外源基因又不是细胞所固有的，和细胞固有基因的表达尚有差别。使用多种宿主细胞，可提高该系统的应用价值。

(3)*转基因系统*(transgenic system)　转基因系统将外源基因整合入动物的生殖细胞，使外源基因在部分或全部组织中表达。该系统和转染系统有一些相同的局限性，即外源基因常以多拷贝存在，整合的位置也和内源性基因不同。

(4)*体外转录系统*(*in vitro* system)　体外转录系统是一种经典的方法。它应用体外转录的方法，结合缺失突变和点突变，来筛选哪些序列是启动子的功能所必需的，哪些

序列对启动子的功能有影响，以及哪些辅助因子对启动子或启动子中的某一片段有何种作用。

启动子研究的第一步是确定启动子的位置及长度。主要方法是用缺失试验来确定启动子的上游边界，即当缺失影响转录始时，说明该处就是启动子的上游边界；用缺失试验结合重组试验来确定下游边界。确定了启动子的位置后，可采用点突变来研究每个碱基在启动子中所起的作用。

研究蛋白质辅助因子与DNA(启动子)的相互作用可采用DNase、足迹法、凝胶阻滞法(band shift)和硫酸二甲酯方法等。

第三节　终　止　子[18~23]

在转录的过程中，提供转录终止信号的序列称为终止子(terminator)。无论是原核生物还是真核生物，通过控制转录的终止来对转录进行调控，也是基因表达调控的重要手段。目前，对原核生物的终止子结构了解较多。

一、原核生物的终止子

通过比较原核生物不同的终止子的序列，发现同源性很差。真正起终止作用的不是DNA序列本身，而是转录生成的RNA。在这一点上，终止子和启动子不同，启动子由DNA序列来提供信号。在大肠杆菌中，有两种类型的终止子，一类为内在终止子(intrinsic terminator)，在体外实验中，只有核心酶和终止子就足以使转录终止，不依赖其他辅助因子；另一类终止子必须在ρ因子(ρ factor)的存在下核心酶才能终止转录，因此称为依赖ρ因子的终止子(ρ-dependent terminator)。第一类终止子又称为不依赖ρ因子的终止子。二种类型的终止子在结构上有一定的差异。

如图5-3所示，不依赖ρ因子的终止子在结构上有两个特征，一是形成一个发夹结构(hairpin)，二是发夹结构末端紧跟着6个连续的U串。不同终止子的发夹结构长度有差异，在7~20 bp之间，由一反向重复序列构成茎，中间的间隔形成环。在茎的底部有一富含GC对的区域。发夹结构中的突变可阻止转录的终止，说明发夹结构的重要作用。经研究确定，新生RNA链的发夹结构可使RNA聚合酶催化的聚合反应暂停，暂停的时间不同终止子有所差异，但典型的终止子暂停时间为60秒左右。转录的终止并不只依赖于发夹结构，新生的RNA中可有多处发夹结构。RNA聚合酶的暂停只是为转录的终止提供了机会，如果没有终止子序列，聚合酶可以继续转录，而不发生转录的终止。6个连续的U串可能为RNA聚合酶与模板的解离提供信号。RNA-DNA间的rU·dA结合力较弱，有利于RNA和DNA的解离。如果将U串缺失或缩短，尽管RNA聚合酶可以发生暂停，但不能使转录终止。DNA上与U串对应的为富含AT对的区域，这说明AT富含区在转录的终止和起始中均起重要的作用。

在依赖ρ因子的终止子中，发夹结构的序列和U串的长度都影响终止子的功能。但在体外实验中，终止子的终止效率变化很大(2%~90%)，没能发现终止效率与终止子结构之间有什么相关性。因此可以肯定还有一些尚未发现的因素在影响着终止子与RNA

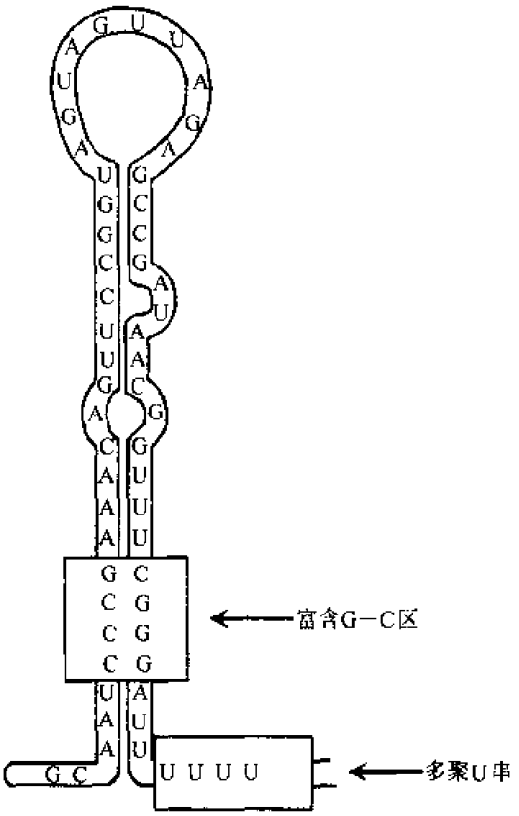


图 5－3　不依赖 ρ 因子的终止子的结构

聚合酶的相互作用。

在体外实验中，人们发现尽管有终止子存在，RNA 聚合酶只在终止子处暂停，但转录并不终止。向该反应系统中加入 ρ 因子，则可使转录在特定的位点终止，产生有独特 3′末端的 RNA 分子。这种终止称为依赖 ρ 因子的转录终止。在大肠杆菌的基因组中，依赖 ρ 因子的终止子相对较少，在噬菌体基因组中较多。经缺失实验表明，ρ 因子可识别终止位点上游 50～90 bp 的区域。分析这段序列的 RNA 发现，C 的含量多，G 的含量少。C 的含量为 41%，而 G 的含量仅为 14%，终止发生在 CUU 中的某一个位置。另外，依赖 ρ 因子的终止子虽有发夹结构，但 GC 含量低，且缺少 U 串。

依赖 ρ 因子的终止子终止位点的上游序列如下：

AUCGCUACCUCAUAUCCGCACCUCCUCAAACGCUACCUCGACCAGAAAGGCG UCUCUU

二、真核生物的终止子

和原核生物相比，对真核生物终止子结构的了解较少。不同的 RNA 聚合酶有不同的终止子。RNA 聚合酶 I 和 III 有类似于原核生物的终止元件，RNA 聚合酶 II 是否有类似的终止元件目前还不十分清楚。RNA 聚合酶 I 的终止子序列为 18 bp，可被辅助因子识别　RNA 聚合酶 III 的终止子类似原核生物，也具有发夹结构和 U 串，发夹结构中有

一段富含 GC 对的序列。另外，还发现一段旋转对称结构，但并不是转录终止所必需。突变分析证明 U 串以外的其他序列也不影响转录的终止。U 串为 4 个，转录常在第三个 U 处终止，也有的在第三或第四个 U 处终止。

第四节　转录的机制[21~32]

转录是由 DNA 指导的 RNA 合成的过程，可分为起始、延伸和终止三个结段。其中，起始和延伸的机制比较复杂，因为对转录的调控都是通过控制转录的起始或终止。也有人将转录分为 4 个阶段，即模板识别、转录起始、延伸和终止。

一、转录的起始

转录的起始是 RNA 聚合酶识别启动子并与之结合从而启动 RNA 合成的过程。这一过程十分复杂，特别是真核生物，需要多种辅助因子参与。原核生物和真核生物在转录起始的过程中有相似也有不同。

1. 原核生物转录的起始

原核生物的 RNA 聚合酶中的 σ 因子起着识别启动子的作用。事实上，在菌体细胞中，只有很少的 RNA 聚合酶分子（核心酶）呈游离状态，大部分与 DNA 分子结合存在。这种结合为松弛型结合，形成松弛型复合物，并且核心酶对启动子无特殊的亲合力。当 σ 因子与核心酶结合形成全酶后，对启动子的亲合力大大提高，形成紧密型复合物。全酶对不同启动子的亲合力是有差别的，最大相差可达 100 倍左右，这也决定了启动子起始转录的效率。全酶要在很长的基因组中（大肠杆菌的基因组长为 4.2×10^6 bp）寻找到约 60 bp 的启动子是一个复杂的过程，目前有三种模式。第一种模式（随机扩散模式）认为全酶与 DNA 呈松弛型结合时，结合和解离的速度很快，这种不断地结合与解离一直持续到全酶寻找到启动子并与之发生紧密型结合。在这种模式中，RNA 聚合酶从一个结合位点到另一个结合位点必须先从第一个位点解离，再与第二个位点结合。因此，RNA 聚合酶移动的速率取决于其扩散速率。理论上推算 RNA 聚合酶与 60 bpDNA 片段间的速率常数的最大值应小于 $10^8 L\cdot mol^{-1}\cdot s^{-1}$，但体外测得的值为 $10^8 L\cdot mol^{-1}\cdot s^{-1}$。如果这一数值适用于体内，随机扩散所需的时间太长，不能解释 RNA 聚合酶与启动子结合的真正方式。所以必然存在另外的途径。如果假设 RNA 聚合酶与基因组 DNA 结合后不再发生解离，可以解决上述矛盾。第二种模式认为 RNA 聚合酶以随机扩散的方式与 DNA 结合后，以一种置换的方式寻找启动子。也就是说与 RNA 聚合酶结合的序列直接被另一段序列代替，酶分子在很快的更换与之结合的序列，直到发现启动子并与启动子形成稳定的复合物。因为解离和结合几乎是同时进行的，所以酶从一个结合位点到另一个结合位点所需的时间很短，使酶分子很快的找到启动子。这种方式使 RNA 聚合酶可以从弱的结合位点很快的移动至强的结合位点。第三种模式认为 RNA 聚合酶分子在 DNA 分子上随机滑动，直到发现启动子并与之结合。这种模式缺少实验证据。就目前所知，还没有一种 DNA 结合蛋白质以这种方式与 DNA 相互作用。

RNA 聚合酶与启动子结合后，与 DNA 形成封闭型启动子复合物，然后 -10 区左右的 DNA 发生熔解，形成 12～17 bp 的单链区，形成开放型启动子复合物，二者均为二元复合物。这时 RNA 聚合酶大约与 75 bp 的 DNA(-55～+20 bp)相互接触。在开放型启动子起始复合物中，RNA 聚合酶上的起始位点和延伸位点被相应的核苷酸前体所占据，在 β 亚基的催化下形成第一个磷酸二酯键，从而形成三元复合物，σ 因子解离下来，形成起始延伸复合物(initial elongation complex)，此时核心酶约与 55 bp(-35～+20 bp)的 DNA 相互接触。

2. 真核生物转录的起始

(1)RNA 聚合酶 I 的转录起始　RNA 聚合酶 I 起始转录需要两种辅助因子 UBF1 和 SL1。UBF1 由一条多肽链组成，可特异地识别核心启动子和 UCE 中富含 GC 对的区域。SL1 单独并没有识别特异 DNA 序列的功能，在 UBF1 和 DNA 结合后，SL1 才可结合上来。只有当两种辅助因子与 DNA 结合后，RNA 聚合酶 I 才能与核心启动子结合，从而起始转录。

SL1 有 4 种蛋白质组成，其中有一种为 TATA 框结合蛋白(TATA binding protein, TBP)，RNA 聚合酶 II 和 III 的起始也需要 TBP 的存在。RNA 聚合酶 I 的 TBP 种间保守性很强，因此它可能不是负责识别特异 DNA 序列的，而是负责与 RNA 聚合酶相互作用的。SL1 类似于原核生物的 σ 因子，它可与启动子特异的结合，并可保证 RNA 聚合酶(包括 I、II 和 III)定位于转录起始位点。

UBF1 与 RNA 聚合酶 I 相互作用可识别不同来源的模板。如鼠的 UBF1 和 RNA 聚合酶 I 可识别人的基因。但 SL1 有种属特异性。

(2)RNA 聚合酶 II 的转录起始　RNA 聚合酶 II 自身不能起始转录，必须在其他辅助因子的作用下才能起始转录。酶和其他辅助因子组成一个基础转录装置(basal transcription apparatus)，可以起始 II 类基因的转录。通用启动子(generic promoter)是指 RNA 聚合酶 II 可起始转录的最短的启动子序列，它可以在任何细胞内表达而无组织特异性。能使这类启动子起始转录所需的辅助因子称为基础转录因子(basal transcription factor)，又称为 TFIIX，X 表示辅助因子的种类。启动子起始转录的效率很低，需要上游的调控序列及相应辅助因子来增强其转录的水平。在真核生物转录起始点左右的序列有一定的规律，转录的第一个碱基多为 A，两边多为嘧啶碱基，可表示为 Py_2APy_5，称为起始子，在 -3 至 +5 之间。起始子是可被 RNA 聚合酶识别的最简单的启动子序列。

转录起始的过程需要很多的辅助因子的参与，它们按一定的顺序与 DNA 结合形成复合物。经足迹法可证明各种辅助因子与 DNA 结合的先后次序。识别 TATA 框及其上游序列的辅助因子称为 TFIID，它由两种组分组成。TATA 框由 TBP 识别。另外还需要其他相关的因子(TBP-associated factors，TAFs)。含不同 TAFs 的 TFIIDs 可识别不同的启动子。RNA 聚合酶 III 的辅助因子 TFIIIB 和 RNA 聚合酶 I 的 SL1 都可看作 TAFs。

转录起始复合物形成的基本过程为 TBP 及 TAFs 先与 DNA 的 TATA 框附近的区域结合，而后依次为 TFIIA、TFIIB、TFIIF 和 RNA 聚合酶、TFIIE、TFIIH 和 TFIIJ，如图 5-4 所示。

TBP 为一种小分子量蛋白(30 kDa)，与 DNA 的小沟结合，这和其他的蛋白因子不

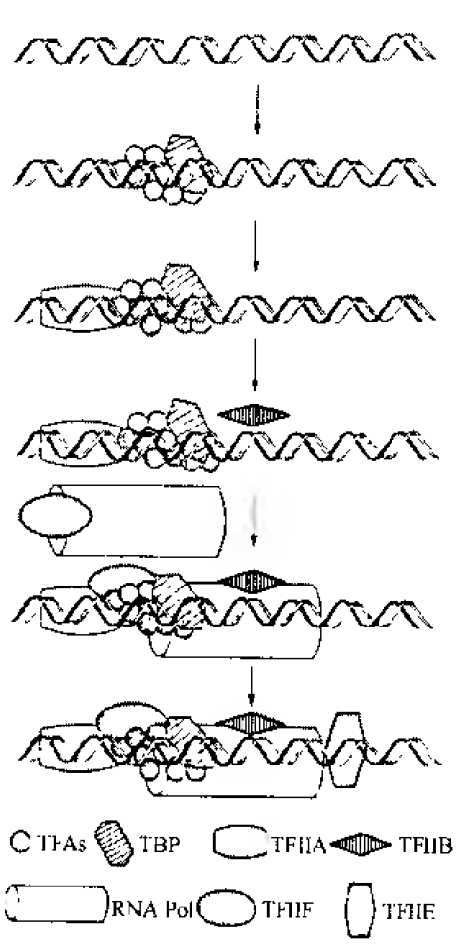


图 5-4 RNA 聚合酶 II 转录起始复合物形成的过程

同，目前所知所有的蛋白因子都是与 DNA 的大沟结合。经分析 TBP 的晶体结构，它作用于 DNA 的一面，形成马鞍状。内面与 DNA 作用，外面与其他的蛋白因子作用。与 DNA 结合的部位的氨基酸序列种间同源性较强。经体外实验测得，TPB 可以保护含 TATA 框的一个螺旋，约在 -37 至 -25 的区域。TBP 是唯一的一个与 DNA 结合具有特异性的基础转录因子。TFIID 具有识别 TATA 框的作用，当 TFIID 结合上来后，形成的复合物可保护 -45 至 -10 的区域。TFIIA 由几个亚基组成，可以激活 TBP，它的结合可使复合物的保护区域扩展。TFIIB 结合上来可使复合物的保护区域扩展至 -10 至 +10，表明该因子与 TATA 框的下游结合。TFIIF 由两个亚基组成，大亚基（RAP76）具有依赖 ATP 的 DNA 螺旋酶的活性，可能参与 DNA 双链的熔解。小亚基（RAP38）与原核生物的 σ 因子在与核心酶结合的部位有同源性，与 RNA 聚合酶 II 紧密结合。TFIIF 的功能是将 RNA 聚合酶 II 与其他的辅助因子组成转录复合物。RNA 聚合酶 II 与复合物结合后可使保护区域在模板链扩展至 +15 处，在非模板链至 +20。上游的保护区域也有所扩展。TFIIF 与复合物结合后可使保护区域扩展至 +30。TFIIH 和 TFIIJ 与复合物的结合并不改变复合物与 DNA 间的作用模式。TFIIH 有激酶的活性，可使 RNA 聚合酶 II CTD 尾磷酸化，结果可使聚合酶与其他的转录起始因子解离，从而得以进行转录的延伸。

RNA 聚合酶 II 的起始过程类似于原核生物，先由 RNA 聚合酶和 DNA 形成封闭型复合物，DNA 解链后形成开放型复合物。在转录起始的过程中，需要上述各种因子及 ATP 的参与。ATP 可能与某些转录因子从复合物上解离有关。TBP 和聚合酶的 CTD 尾间的反应对转录起始有重要意义。具体的机制尚待进一步探讨。

在不含 TATA 框的启动子的转录起始中，同样需要多种转录因子的参与。其中，多数资料表明 TFIID 可能具有识别起始子的功能，并与之特异结合，但 TFIID 与起始子的结合要比和 TATA 框的弱。此外，TFIID 与 DNA 还有其他的作用方式。研究表明 TFIID 与起始子的结合可能与 TFII-I 及 RNA 聚合酶 II 有关。TFII-I 也是一种可与起始子结合的蛋白因子，它对起始子起始转录的功能十分重要，但具体的机制还不十分清楚。TFII-I 与碱性蛋白（basic helix-loop-helix，bHLP）USF 蛋白在免疫学上有交叉反应，表明是 USF 的一种。另外还发现 YY1 具有识别起始子的功能。YY1 含有锌指结构，与含有 CCAT 的序列有亲和力。

参与转录起始过程有多种转录因子，很多转录因子又有多亚基组成，共计有 20 多种多肽分子；RNA 聚合酶本身有 10 多个亚基组成，所以形成的起始复合物十分庞大，总计在 1000 kDa 以上。

在此类启动子的结构中，还有其他的元件如 CAAT 框、GC 框和八聚体元件。对各种元件的识别都需要特有的转录因子。如 SP1 可识别 GC 框。GC 框在不同的启动子中位置差异很大。SP1 由一个亚基组成，分子量为 105 kDa，可结合 6 bp 的区域。CAAT 框由 CTF 家族的因子识别。CTF 家族由一个基因经不同的加工方式而形成，对 CAAT 有相同的亲合力。其他识别 CAAT 框的因子有不同的特点，如 CP1 对 α-球蛋白及晚期腺病毒的 CAAT 框亲和力高；CP2 对 γ-纤维蛋白基因的亲合力较高；CP3 也有不同的特异性。识别 CAAT 框还有其他的蛋白因子，如鼠肝中有两种蛋白因子，C/EBF 可识别 GCAAT，ACF 可识别 CCAAT。识别八聚体元件的也不止一种蛋白因子，在非淋巴细胞中为 Oct-1，在淋巴细胞中为 Oct-2。Oct-2 可促进免疫球蛋白 κ 基因的转录。至于这些元件的识别与转录起始复合物形成的关系，目前尚不十分清楚。

(3)RNA 聚合酶 III 转录的起始　被 RNA 聚合酶 III 转录基因的启动子有三类，其中有二类属于内部启动子。这两类启动子转录的起始需三种辅助因子 TFIIIA、TFIIIB 和 TFIIIC 的参与。TFIIIA 已被克隆，是一种含锌指结构的蛋白。TFIIIB 由 TBP 和其他两种蛋白组成。对 TFIIIC 了解不是十分清楚，其分子量很大(大于 500 kDa)，至少由 5 个亚基组成。如图 5-5 所示，两种启动子的起始过程有所区别。

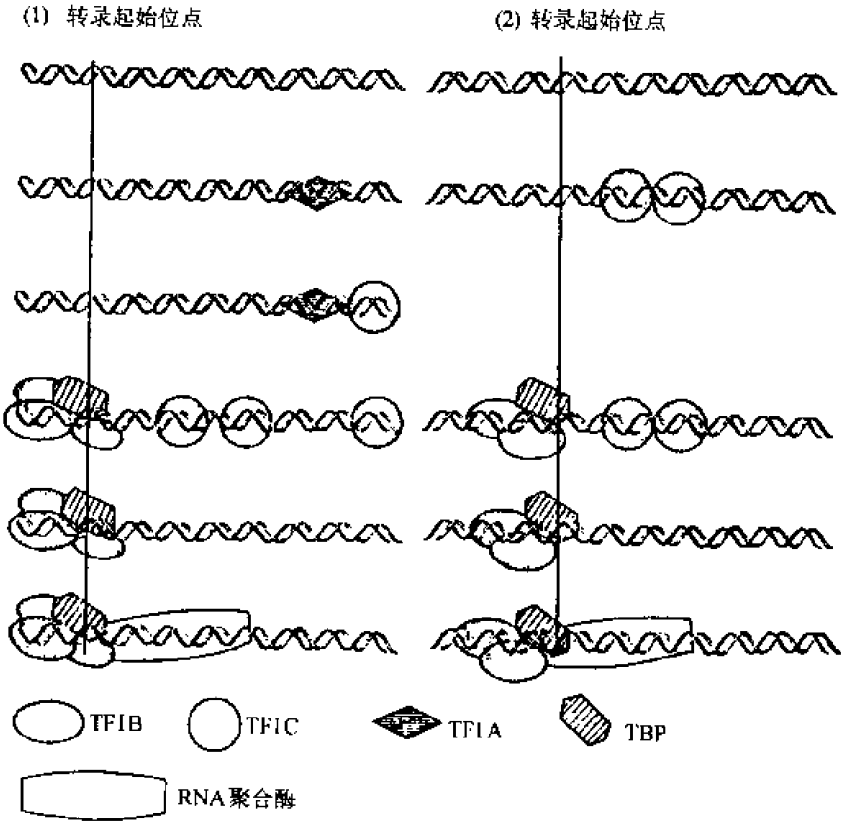


图 5-5　两种启动子转录起始复合物的装配过程

在第二类内部启动子，TFIIIC 识别 B 框，但可以与包括 A 盒在内的广泛区域结合。然后 TFIIIB 与起始位点附近的区域结合。在第一类内部启动子，TFIIIA 与包括 C 盒在内的序列相结合，而后使得 TFIIIC 能与靠近下游的序列结合，最后 TFIIIB 才能与转录起始点附近的序列结合，并且伴随着另一个 TFIIIC 与 DNA 的结合。经上述的复合物形成后 RNA 聚合酶即可与启动子结合。RNA 聚合酶 III 加入起始复合物中后，TFIIIC 和 TFIIIA 都要释放下来。现已确定了 TFIIIB 的两种蛋白质分子与 −40 至 +11 的区域结合。

在高盐条件下，TFIIIA 和 TFIIIC 可以从启动子上解离下来而不影响转录的起始。在这种条件下，TFIIIB 依然可以与转录起始点附近的序列结合，并且保证 RNA 聚合酶 III 与启动子结合。因此，TFIIIB 是 RNA 聚合酶 III 所需要的真正转录因子；而 TFIIIA 和 TFIIIC 属于装配因子(assembly factor)，起协助 TFIIIB 与启动子正确结合的作用。TFIIIB 的主要作用是使 RNA 聚合酶 III 正确定位于启动子上。从其功能上看，类似于 SL1 及原核生物的 σ 因子，自身缺乏与 DNA 结合的能力，但有与其他蛋白因子协同作用的功能。TFIIIB 也含有 TBP。

RNA 聚合酶 III 的第三类启动子(snRNA 的启动子)不是内部启动子，其结构类似于 RNA 聚合酶 II 的启动子。TATA 框的识别同样依赖于 TBP，但必须有其他 RNA 聚合酶 III 的辅助因子共同作用，使 RNA 聚合酶 III 在启动子上正确定位。这也可以解释为什么这类启动子须由 RNA 聚合酶 III 来启动转录。

TBP 在三种 RNA 聚合酶的转录起始中均有相同的作用。从对酵母和动物转录的研究中得出，所有 RNA 聚合酶 III 的转录起始包括内部启动子都需要 TBP；在 RNA 聚合酶 I 识别 rDNA 的启动子并起始转录的过程中，TBP 是 SL1 的组分，起定位 RNA 聚合酶 I 的作用；在 RNA 聚合酶 II 的转录起始过程中，TBP 可以识别 TATA 框，同样起定位 RNA 聚合酶的作用。因此，尽管 TBP 在不同的启动子中与其他转录因子形成复合物的方式不同，但功能都是一样的，即起定位作用。所以 TBP 又称为定位因子(positioning factor)。

RNA 聚合酶 III 看来并没有对启动子特异序列的识别作用。它能正确地结合在转录起始点附近，全靠其他转录因子的作用。在二类内部启动子中，TFIIIA、TFIIIC 和 TFIIIB 的装配使 TFIIIB 结合在转录起始点附近，RNA 聚合酶 III 与 TFIIIB 的特异结合保证了酶分子的正确定位；在上游启动子中，同样是靠 TBP 的作用。

二、转录的延伸

转录的起始第一个涉及的问题是第一个碱基的选择。在原核生物，起始转录的碱基距保守 T 为 6 至 9 个核苷酸，多为 CAT 模式，第一个碱基为 A。这可能是由于 RNA 聚合酶的空间结构决定的，即结合部位与聚合活性部位间的距离决定了转录起始碱基与 TATA 框(结合位点)的距离。

在原核生物的 RNA 聚合酶中，催化 RNA 合成的可能是 β 亚基。在酶分子上有两个核苷酸结合位点，一是起始核苷酸位点，一是延伸核苷酸位点。当两个位点都与模板互补的核苷酸结合后，第一个磷酸二酯键才能形成。

RNA 聚合酶在转录起始后，对 DNA 的亲合力会发生变化。对于原核生物，核心酶本

来对 DNA 有一定的亲合力，但全酶对启动子有很高的亲合力。转录起始后，σ 因子解离下来，但这时的核心酶就既能与 DNA 结合，又能在 DNA 分子上滑动。真核生物是靠多种辅助因子的共同作用下来维持这一状态的。

延伸过程中的另一问题是拓扑学问题。在 DNA 解链过程中，位于 RNA 聚合酶前面的双链 DNA（未被转录的）会越缠越紧，产生正超螺旋；后面的则越来越松，产生负超螺旋。为了消除这种现象，需要旋转酶（gyrase）和拓扑异构酶 I（topoisomerase I），前者可消除负超螺旋，后者可消除正超螺旋。在酵母细胞中，如果缺乏消除负超螺旋的酶，在转录区域的负超螺旋会成倍的增加。另外，转录本身受 DNA 超螺旋状态的影响。多数情况下，DNA 的超螺旋状态有利于转录。但也有时超螺旋状态抑制转录，目前还无法解释。

新生的 RNA 链与 DNA 形成的杂交双链很短，只有几个碱基，有人提出只有 2 个。经胰 RNA 酶（只切断单链 RNA）消化转录系统时，得到的 RNA 片段为 10 个碱基左右。这一数值不能准确表示 RNA 与 DNA 形成杂交链的长度，因为部分 RNA 可能受到了 RNA 聚合酶的保护。

RNA 合成的速度约为每秒 40 个核苷酸，与蛋白质合成的速度相近（15 个氨基酸/秒），但比 DNA 复制的速度（800 bp/秒）要慢得多。RNA 聚合酶在 DNA 分子上的运动不是匀速的，在经过富含 GC 对的序列 8 至 10 个核苷酸后，会发生一次暂停，这与转录的终止有关。

三、转录的终止

转录进行到终止子序列时，就进入了终止阶段，包括新生 RNA 链的释放及 RNA 聚合酶与 DNA 解离。

1. 转录终止的机制

前已提及，原核生物的终止子有两种，一种是不依赖 ρ 因子的终止子，一种是依赖 ρ 因子的终止子。在不依赖 ρ 因子的终止子中，转录的终止在没有 ρ 因子的情况下依靠终止子本身的结构就足以使转录终止，如 U-dA 间的结合较弱，为 RNA-DNA 杂交体的分离提供了条件。在依赖 ρ 的终止子中，必须在 ρ 因子的存在下，转录才能终止。

ρ 因子的分子量为 46 kDa，活性形式为六聚体，分子量为 270 kDa。体外实验表明，当 ρ 因子与 RNA 聚合酶的浓度比为 1∶10 时，表现出最大的活性。ρ 因子有依赖 RNA 的 ATPase 活性，RNA 的长度要大于 50nt，这说明它需要与 RNA 结合。ρ 因子可以与新生 RNA 在终止子上游的某一处与 RNA 结合，这种结合可能需要某种特异的序列。也可以与 RNA 的 5′端结合。ρ 因子要追赶上 RNA 聚合酶，才能发挥其终止作用。至于 ρ 因子是如何赶上 RNA 聚合酶的，有人认为 ρ 因子沿 RNA 链移动的速度比 RNA 聚合酶在 DNA 链上的移动速度快，聚合酶在终止子处的暂停给 ρ 因子提供了机会。经实验证明，在无 RNA 聚合酶的存在下，ρ 因子具有使 RNA-DNA 杂交体解离的作用，这一作用需要 ATP。因此推测 ρ 因子可使转录泡的 RNA-DNA 解链，然后 ρ 因子及 RNA 聚合酶释放，转录终止（图 5-6）。

在原核生物中，转录与翻译是同时进行的，ρ 因子的终止作用可被核糖体的存在而阻

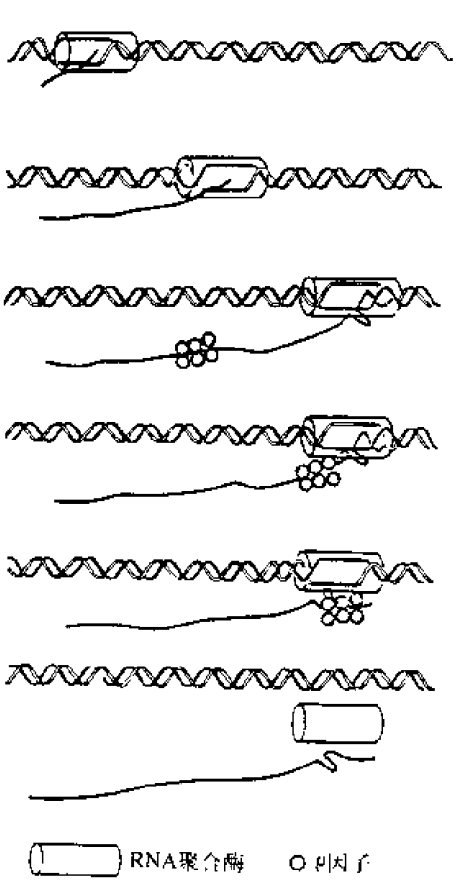


图 5-6 ρ因子的作用模型

碍。当生长旺盛时，mRNA 被多个核糖体所结合，ρ因子没有机会与 mRNA 结合。即使结合上去，也会受到核糖体的阻碍而不能实现终止作用。这样为原核生物基因表达的调控提供了一种方式。上述模式可解释下列实验现象。在一个转录单位内，如发生无义突变，可阻止转录单位突变位置之后序列（远端序列）的转录。在正常情况下，当 RNA 聚合酶转录至终止子处时，由于核糖体与 mRNA 结合，ρ因子不能实现终止作用。无义突变使得翻译提前终止，核糖体提前释放，ρ因子就有机会追赶 RNA 聚合酶，从而引起转录的终止。ρ因子的突变可阻止这种现象的出现，也就是说，ρ因子的缺乏会使转录继续进行。另外，RNA 分子中的二级结构也可阻止 ρ因子的作用。

ρ因子的突变对转录终止的影响变化很大。体外实验证明，不同的依赖 ρ因子的终止子对 ρ因子浓度的要求高低不一。因此，在 ρ因子的渗漏突变时，不同终止子的反应也有所区别。ρ因子的突变可被其他的基因突变所抑制。在 ρ因子突变引起的转录不能终止的菌株中，RNA 聚合酶 β 亚基基因（*rpoB*）的一种突变可以恢复转录的终止。*rpoB* 的另一种突变可减弱依赖 ρ因子的转录终止，说明 β 亚基可能是 ρ因子的作用部位。

对真核生物转录终止的机制还不十分清楚。在真核生物中同样存在终止子的结构。在 RNA 聚合酶 I 的终止过程中，需要一辅助因子识别一 18 bp 的终止子序列。RNA 聚合酶 II 和 III 的终止子与原核生物不依赖 ρ的终止子结构类似，可能也具有类似的终止机制，终止可能靠 RNA 聚合酶本身来完成，而不像转录的起始那样需要多种辅助因子。终止子中有对称结构，但这不是终止所必需的。RNA 聚合酶 II 的终止子序列与 RNA 聚合酶 III 相似，RNA 聚合酶 II 对终止序列的识别特异性差一些，终止可发生在广泛的区域，而缺少特定的识别点。在有些转录单位中，终止位点位于 mRNA 成熟 3′端下游大于 1000 bp 处。由于真核生物成熟 mRNA 的 3′已经被切除一段且添加了 poly(A)，所以不能确定转录终止子处的序列。在 RNA 聚合酶 III 的终止子中，单靠 U 串（4 个 U）是不能终止转录的，其富含 GC 的上下文的序列也起重要作用，终止常发生在 U 串的第二个 U。相对原核生物，我们对真核生物转录终止的了解还不充分。

2. 抗终止作用

抗终止作用（antitermitation）是细菌操纵子和噬菌体基因表达调控的一种方式，首次发现于噬菌体感染过程中。噬菌体只有少数基因（早期基因）能被细菌 RNA 聚合酶转录，而这些基因的产物能起始其他启动子的转录，或使 RNA 聚合酶阅读通过终止子，从

而调控其他基因的表达。

改变识别启动子的一种机制是更替 σ 因子；另一种途径是合成新的噬菌体 RNA 聚合酶。不论是哪种情况，都是使转录由新的启动子开始，而早期基因的转录终止。噬菌体的早期基因和中、晚期基因相邻，中间由终止子隔开。如果转录不能在终止子处终止，中、晚期基因将会表达。也就是说中、晚期基因和早期基因使用同一个启动子。因此，抗终止作用不是更替启动子，而是转录的进一步延伸。

抗终止作用首次发现于 λ 噬菌体。宿主 RNA 聚合酶首先转录两个基因，称为快早期基因，在抗终止作用下，转而转录迟早期基因。这种转变受 *N* 基因的抑制。从遗传学的角度来看，重新识别启动子和抗终止作用都属于负性调控，早期基因的表达是其他基因表达的先决条件。

在 λ 噬菌体中，快早期基因有两个转录单位，分别从 P_L 和 P_R 启动子开始转录。大肠杆菌的 RNA 聚合酶分别终止于终止子 t_{L1} 和 t_{R1}，二者均是依赖 ρ 因子的。*N* 基因的表达产物 pN 为终止因子，可使 RNA 聚合酶通过终止子，转录后面的基因。晚期基因编码装配噬菌体的组分，它的表达受 *Q* 基因的调控。*Q* 基因属于迟晚期基因，编码的产物为 pQ，也是一种抗终止蛋白，使 RNA 聚合酶由 P_R 开始转录。这样，抗终止因子就调控了不同基因的表达。

pN 的抗终止活性高度特异，其抗终止作用不依赖于终止子。在转录单位的上游有 pN 识别的特异序列，称为 *nut*。这也说明蛋白质与 DNA 间的识别位点和作用位点并不是一致的，是可以分离的。控制左向转录的 *nut* 位点称为 *nutL*，控制右向转录的称为 *nutR*。*nutL* 位于 P_L 启动子的起始位点和 N 编码区域之间；*nutR* 位于 *cro* 基因和 t_{L1} 之间。这表明两个 *nut* 位点位于不同的区域，*nutL* 接近于终止子；*nutR* 接近于启动子。抗终止作用是实现并不参与转录的起始和终止过程。不同的噬菌体有不同的 *N* 基因、*nut* 位点和 pN。

第五节　转录产物的后加工[33~38]

多数转录的初始产物并无生物学活性，必须经进一步的加工处理才有生物学活性，特别是对于真核生物转录产物的后加工更为重要。RNA 的剪接也是后加工的重要内容，但内容较多，将在下一节介绍。

一、原核生物转录产物的后加工

在原核生物中，mRNA 的寿命非常短，有的或大多数半衰期只有几分钟，这是原核生物基因表达调控的一种手段。而 rRNA 和 tRNA 比较稳定，半衰期为几个小时。通过比较原核生物成熟的 rRNA 和 tRNA 与其转录产物，可以发现：两种 RNA 成熟分子的 5′端为单磷酸，而原始转录产物为三磷酸；成熟分子比原始转录产物小；成熟分子含有异常碱基，而原始转录产物中没有。因此，rRNA 和 tRNA 的转录产物必然存在着后加工过程。

1. rRNA 的后加工

在原核生物中，有三种 rRNA(5 S、16 S 和 23 S)。其基因(rDNA)与 tRNA 的基因

(tDNA)混在一起排列在一个操纵子(*rrn*)中。大肠杆菌中有7个这样的操纵子。各种基因在操纵子中的排列没有一定的规律,种类、数量和位置都不相同。但三种rRNA的相对位置有一定规律,一般为16 S rDNA、tDNA、23 S rDNA、5 S rDNA,最后为tDNA。在16 S rRNA和23 S rRNA之间的tDNA有400~500 bp。在7个 *rrn* 操纵子中,有4个在16 S rRNA和23 S rRNA间的tDNA为 $tRNA_2^{Glu}$,其他的三个操纵子为 $tRNA_1^{Ile}$ 和 $tRNA_1^{Ala}$。在有些 *rrn* 操纵子中,5 S rDNA后没有其他的基因序列,而另一些 *rrn* 操纵子中有tRNA的基因,如 *rrnC* 中为 $tRNA^{Trp}$ 和 $tRNA_1^{Asp}$。大肠杆菌中的7个操纵子多数位于复制起始点的两侧,且转录方向与DNA的复制方向一致。*rrn* 操纵子有两个启动子,第一个为 *p*1,位于16 S rRNA序列上游300 bp左右,可能为主要的启动子。第二个启动子 *p*2 位于 *p*1 下游110 bp左右。该操纵子转录的产物为30 S RNA。RNA酶III负责该转录产物的后加工。在缺乏RNA酶III的菌体细胞中,没有成熟rRNA的出现,且30 S RNA产生堆积。该酶在体外能作用于30 S RNA,产生p16和p25。p16和p25比成熟的产物只稍长一些,可进一步转变为成熟的rRNA。在p16和p25 RNA的初始转录产物中,5′和3′端的区域碱基互补,形成双链结构。互补区域的间隔序列相对稳定,p16为1600 nt,p23为2900 nt。RNA酶III在p16和p23的酶切位点无明显的同源性。RNA酶III除要求切点或其周围为双链结构外,还不清楚是否有其他的条件。它可能取决于一级结构和二级结构的共同作用。

经RNA酶III作用后的产物需经进一步的转变才能成为成熟的分子。其具体的机制尚不十分清楚。

2. tRNA的转录后加工

原核生物的tRNA的初始转录产物可以包含几个同种tRNA的基因拷贝,也可以含有几个不同种的tRNA。还有些tRNA与rRNA转录在一起。

tRNA的初始转录产物要经过多种加工处理后才能变为有功能的分子。目前了解较多的是大肠杆菌的 $tRNA_1^{Tyr}$。它由85个核苷酸组成。其编码基因有两个基因拷贝,长350 bp。两个基因拷贝间有200 bp的间隔子。

有两种类型的tRNA基因,一种具有CCA序列,称为I型;另一种没有CCA序列,称为II型。I型tRNA的CCA下游仍有一段序列,需在酶的作用下切除。II型tRNA需在酶的作用下添加CCA序列。原核生物的tRNA多为I型,少数噬菌体如T4为II型。

tRNA的后加工是在多种酶的作用下完成的,目前这方面的研究进展较多。下面简要介绍几种参与tRNA后加工的酶。

(1)RNA酶P　RNA酶P(RNase P)是一种内切酶,能使tRNA操纵子和rRNA操纵子转录产物中的tRNA产生成熟的5′末端。该酶由两个组分组成,一是20 kDa的蛋白质组分(C_5 蛋白),一是由375个核苷酸组成的120 kDa的RNA组分(M_1RNA)。它们以紧密结合的核蛋白形式存在。这两种组分高度保守,大肠杆菌的RNA酶P成分可与人类的RNA酶P成分能交叉重建为有活性的RNA酶P。各种成熟的tRNA在5′端序列没有同源性,但RNA酶P均能使之产生正确的5′末端。因此,RNA酶P识别的是tRNA的空间构象,特别是5′端的构象,并不识别切点附近的序列。在不同的tRNA分子中也存在少数保守核苷酸,可能是维持空间构象所必需的。该酶的专一性很强,只作用于tRNA前

体，但无种属特异性。

(2)RNA 酶 D　　RNA 酶 D 为核酸外切酶，能逐个切除 tRNA 3′端多余的核苷酸，使 CCA 暴露。体外实验证明，该酶对 CCA 下游序列呈高活性，对 CCA 上游序列也有一定的活性。在体内，当 CCA 末端产生后，会立即产生氨酰化而得到保护。RNA 酶 P 存在时 RNA 酶 D 可达最大活性。另外，还有一些与该酶作用相同的酶，如 RNA 酶 Q、RNA 酶 Y 和 RNA 酶 P_3 等。

(3)tRNA 核苷酸转移酶(tRNA nucleotidyl transferase)　　该酶能以 ATP 和 CTP 为前体，催化 tRNA 的 3′端生成 CCA。尽管大肠杆菌的 tRNA 大多为 I 型，但编码 tRNA 核苷酸转移酶的基因(*cca*)突变后，会影响菌体的生长。这是因为 tRNA 的 CCA 末端常处于核酸酶的降解作用下而发生丢失。在这种情况下，该酶就起到了修复作用。该酶只需要 ATP、CTP 和 tRNA，就能生成 CCA。另外，该酶几乎没有种属特异性。实验证明，该酶识别 tRNA 的空间结构。

(4)RNA 酶 III　　RNA 酶 III 和 RNA 酶 O、RNA 酶 P_2 的功能相似，负责切开间隔序列。

另外，还有其他的一些酶类，如 RNA 酶 P_4，参与 tRNA 的后加工。

原核生物的 mRNA 不需要进一步的加工处理就可以具有生物学活性，做为翻译的模板。在最近的研究中，发现原核生物的 RNA 转录后也有在 3′添加 poly(A)的现象。大肠杆菌的 poly(A)聚合酶早在 1962 年就已发现。以后发现了真核生物的 mRNA 具有 3′的 poly(A)的结构，人们一直认为它是真核生物 mRNA 的特征。但后来发现原核生物的 RNA 也存在 3′添加 poly(A)的现象。在质粒拷贝数 *B*(plasmid copy number B, *pcnB*)基因的缺失突变(*pcnB*　)的研究中，发现缺失突变体对 RNA I 的降解非常慢。从突变体中分离到的 RNA I(结构类似 tRNA)缺乏 3′poly(A)的结构，而在非突变体($pcnB^+$)中，RNA I 具有 poly(A)的结构，却降解的非常快。因此，RNA I 的 3′腺苷化使其易于被降解，这和真核生物的 mRNA 有异曲同工之处。同样，在 $pcnB^-$ 中，一些 mRNA 的降解速度也变慢，且 RNA I poly(A)的长度变长。RNAI3′端的降解主要由 PNPase 来负责，5′端的降解主要由 RNA 酶 E 来负责。3′端的 poly(A)可能为 PNPase 提供特定的信号，使其对 RNAI 降解加快。尽管对这种生物学现象还缺乏了解，但它必定有其存在的意义，并且还可能发现一些相类似的现象。

二、真核生物的转录后加工

真核生物转录后加工除 tRNA 和 rRNA 外，mRNA 也要经过后加工的过程。有关真核生物 tRNA 的加工处理了解还不多，可能和原核生物有相似之处。下面介绍 rRNA 和 mRNA 的后加工

1. rRNA 的转录后加工

真核生物有 4 种 rRNA，即 5.8 S rRNA、18 S rRNA、28 S rRNA 和 5 S rRNA。其中，前三者的基因组成一个转录单位，产生 47 S 的前体，并很快转变成 45 S 前体。真核生物的 rRNA 的成熟过程比较缓慢，所以其加工的中间体易于从各种细胞中分离得到，使

得对其加工过程也易于了解。哺乳动物的45S前体包含着 18 S、28 S 和 5.8 S rRNA，但其长度是三种成熟 rRNA 长度和的二倍。共有 110 多个甲基化位点，在转录过程中或以后被甲基化。甲基基团主要是加在核糖上。这些甲基化位点在加工后仍保留在成熟的 rRNA 中，其中，18 S rRNA 上有 39 个，另外还有 4 个是在进入细胞质后甲基化的；74 个在 28 S rRNA 上。这表明甲基化是 45 S 前体上最终成为成熟 rRNA 区域的标志。

由 45 S 前体加工成成熟的 rRNA 有两种方式，一种发现于 HeLa 细胞，一种发现于鼠的 L 细胞。两种方式中 45 S 前体的剪切位点是相同的，只是对剪切位点的剪切顺序不同。剪切位点一个位于 18 S rRNA 的 5′边，两个位于 18 S rRNA 和 5.8 S rRNA 间的间隔区，另两个位于5.8S rRNA 和 28 S rRNA 间的间隔区。也可能存在另外的方式。有时在一种细胞中可以发现两种以上的成熟方式。目前还不清楚在剪切位点断裂后是否就产生成熟的末端，还是要经进一步的加工。对负责加工的酶类也知之不多。但可肯定的说，加工过程需要蛋白质的参与，可能形成核蛋白体的形式。

真核生物的 5 S rRNA 也是和 tRNA 转录在一起的，经加工处理后成为成熟的 5 S rRNA。

2．mRNA 的后加工

在真核生物中，几乎所有的成熟 mRNA 有 5′帽子(cap)结构，多数还有 3′的 poly(A)尾。这些结构都是在转录后经过修饰的结果。

成熟真核生物的 mRNA 并没有游离的 5′端，而是一种被称为帽子的结构。不同真核生物的 mRNA 可有不同的帽子结构，同一种真核生物的 mRNA 也常有不同的帽子结构。同一种 mRNA 是否有不同的帽子结构，目前尚不清楚。真核生物的帽子结构可归纳为三种：m7GpppX 为帽子 0；m7GpppXm 为帽 1；m7GpppXmYm 为帽 3。

帽子结构的作用还不十分清楚，但可以肯定具有重要的功能。目前认为它有下列作用：①为核糖体识别 mRNA 提供信号；②增加 mRNA 的稳定性；③与某些 RNA 病毒的正链 RNA 的合成有关。

多数真核生物(酵母除外)的 mRNA 3′具有约为 200 bp 长的 poly(A)尾巴。具有此特征的 mRNA 表示为 $poly(A)^+$，不具有该特征的写为 poly(A)$^-$。poly(A)尾巴不是由 DNA 所编码，而是在转录后由 RNA 末端腺苷酸转移酶(RNA terminal riboadenylate transferase)催化下，以 ATP 为前体，添加到 mRNA 的 3′末端。核 RNA 也具有 poly(A)结构。无论是 mRNA 还是核 RNA，poly(A)与 poly(A)结合蛋白[poly(A)-binding protein，PABP]结合。PABP 的单体分子量为 70 kDa，可与 poly(A)序列中的 10～30 个碱基结合。RNA 末端腺苷酸转移酶又称为 poly(A)聚合酶，分子量为 300 kDa。poly(A)的添加位点并不在 RNA 转录终止的 3′端，而是首先由一 360 kDa 的内切酶和其他因子识别切点上游 13～20 个碱基的 AAUAAA 和下游的 GUGUGUG(有些情况例外)，然后切除一段序列，在此基础上，由 RNA 末端腺苷酸转移酶催化添加 poly(A)。AAUAAA 保守性很强，只有少数情况下有一个碱基的差异，这段序列的突变可阻止 poly(A)的形成。对 poly(A)形成的机制可通过体外添加实验来加以研究。识别 AAUAAA 的特异因子由 3 个亚基组成，每个亚基均有和 RNA 非特异结合的能力，但单独均没有与 AAUAAA 特异结合的能力。三个亚基间的相互作用可能使其具备了特异识别 AAUAAA 的能力。特异因子与内

切酶及 poly(A)聚合酶形成复合物存在，由这一复合物先完成剪切过程，紧接着进行 poly(A)的添加。poly(A)聚合酶单独的催化作用没有特异性，与其他的两个成分结合后才有特异性。poly(A)聚合酶催化 poly(A)的生成可分为两个阶段，首先在 mRNA 的 3′添加 10 个左右的 poly(A)，该过程严格依赖 AAUAAA 的存在，且是在特异因子的协助下完成的。第二个阶段将 poly(A)尾延长致全长，这一过程不依赖 AAUAAA，而是需要另外一个刺激因子识别已形成的 poly(A)并引导其延长。

有些 mRNA 的 3′端无 poly(A)尾，如组蛋白的 mRNA。其 3′末端的正确形成依赖于 RNA 本身形成的茎环结构，即转录终止于此处。对茎环结构进行突变，即便是改变了序列，只要维持其二级结构，就不影响转录的终止，这说明二级结构所提供的信息比序列本身更重要。其成熟 3′末端的形成也需要剪切过程。如组蛋白 H3 mRNA，转录原始产物要经切除一段序列才能成为成熟的 mRNA。有三个成分参与了该过程，第一种是热敏因子，功能不详；第二种因子识别剪切位点上游的发夹结构；第三种是 U7 snRNA，长 56 个碱基，其 5′末端一保守序列能与 mRNA 的剪切点下游的 GAAAGA 互补（如图 5－7 所示）。U7 snRNA 的序列和结构在人类和海胆有很高的保守性。

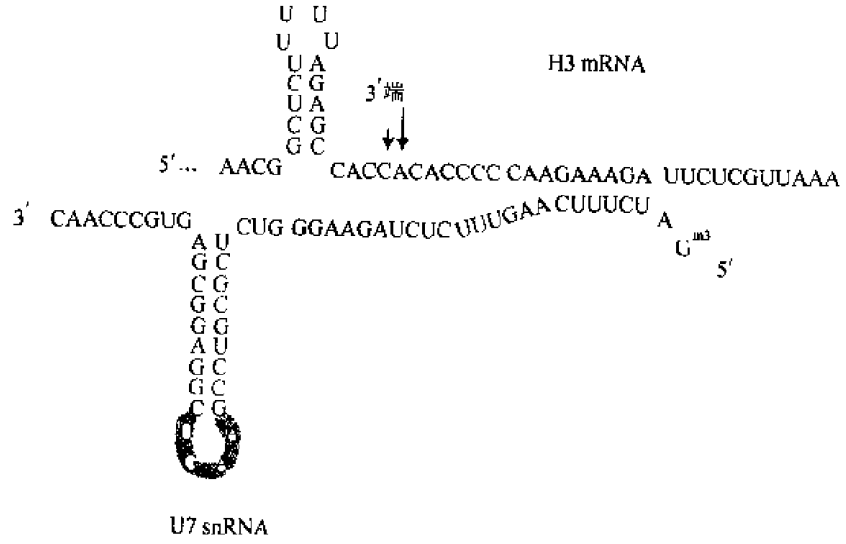


图 5－7 组蛋白 mRNA3′成熟末端的形成

另外，还有其他一些成分参与了 poly(A)的形成过程，如 U1 snRNA(small nuclear RNA)。在海胆中，还发现一种 12 S 的因子参与多聚腺苷酸化，分子量为 250 kDa，含有一 60 个核苷酸的 RNA 组分。现已明确，poly(A)在细胞质中可受核酸酶的降解，但总是维持在 100～200 bp，其机制尚不十分清楚，但细胞质中有一种 poly(A)聚合酶，能延长已有的 poly(A)。关于 poly(A)的功能，目前认为与 mRNA 的寿命有关。如去掉 poly(A)的 mRNA 易被降解。但 poly(A)的长度与其寿命间没有发现相关性。poly(A)与 poly(A)结合蛋白的缺乏能阻止酶的水解。poly(A)的缺失可抑制体外翻译的起始。PABP 的缺乏能抑制酵母的翻译。在胚胎发育过程中，mRNA 的 poly(A)化对其翻译有影响。非 poly(A)化的 mRNA 作为一种贮存形式，在 poly(A)化后开始翻译。对含 poly(A)的 mRNA 去 poly(A)，可减弱其翻译。

poly(A)在分子生物学实验中有很大的应用价值。应用mRNA的该特性，可用寡聚(oligo)(dT)为引物，反转录合成cDNA；也可将oligo(dT)与载体相连，用于从总RNA中分离纯化mRNA。

3. 后加工与mRNA的稳定性

转录后加工是产生各种成熟RNA分子的重要过程。如原核生物的rRNA和tRNA，真核生物的rRNA、tRNA和mRNA都要经过后加工的过程。各种RNA分子的加工都是在特定的酶参与下完成的，包括核酸内切酶和核酸外切酶。核酸酶除参与后加工过程外，还负责过剩RNA的降解。特别是mRNA，一直处于被降解的过程中。无论是原核还是真核生物，各种RNA分子在细胞内转录的强弱是不同的。被降解的速度也不同，但转录和降解可以达到一种平衡状态，使细胞内各种mRNA分子保持在特定的均衡水平。每种核酸酶对其水解的靶序列有一定的要求。但目前还没有弄清哪些特殊的序列在决定着RNA的稳定性。

原核生物mRNA降解的方向为5′至3′，可能是由内切酶在最后一个核糖体后面发挥作用。余下的片段由外切酶按3′至5′水解成核苷酸。现在已经发现了几个内切酶水解mRNA的位点。在核酸酶E缺失株中，mRNA和小核糖体RNA的稳定性提高2－3倍。说明核酸酶E负责mRNA和小核糖体RNA的降解。由内切酶作用所产生的片段对外切酶的敏感性不同。mRNA分子中的二级结构可以阻止外切酶的作用，对其5′端的序列产生保护作用。T4噬菌体的mRNA稳定性较高，且该mRNA的5′序列与细菌mRNA结合后也可提高细菌mRNA的稳定性。因此，内切酶的酶切位点的有无并不是决定mRNA稳定性的唯一因素。

真核生物mRNA的稳定性可能取决于去稳定元件(destablizing elemant)。将这一元件与mRNA结合可引起该mRNA的降解。将一个去稳定元件从mRNA分子中去掉，并不能增强其稳定性，说明可能在一个mRNA分子中有多个去稳定元件。目前已发现了两种类型的去稳定元件。

在有些不稳定mRNA分子中有一共同的结构特征，即3′端区域有一约50碱基的富含AU的序列，称为AU元件(AU-rich element，ARE)，其保守序列为AUUUA，重复7次。ARE可以诱导mRNA脱去poly(A)，失去了PABP的保护，易被降解。在转铁蛋白的mRNA有一称为IRE的序列，位于3′端非翻译区，控制着mRNA对铁离子浓度变化的关系。IRE含有茎环结构，与一种蛋白质结合，该蛋白与mRNA的结合力受铁离子的控制。当铁离子与蛋白质结合后，可与IRE结合，从而抑制了去稳定序列的作用；当铁离子与蛋白质解离后，不能与IRE结合，mRNA被降解。

参与RNA降解的核酸酶类可能比参与RNA成熟过程的核酸酶特异性差一些。特异性较高的核酶主要识别RNA的空间结构，如特殊的二级或三级结构，而不是一段特异的序列。

第六节　RNA的剪接[39～59]

割裂基因(split gene)即不连续基因(interrupted gene)，是指编码某一RNA的基因中

有些序列并不出现在成熟 RNA 的序列中，成熟 RNA 的序列在基因中被其他的序列隔开，这些序列称为介入序列，又称内含子(intron)。被内含子隔开的出现在成熟 RNA 中的序列称为外显子(exon)。一个基因的外显子和内含子共同转录在一条转录产物中，然后将内含子去除而把外显子连接起来形成成熟的 RNA 分子，这一过程称为 RNA 剪接(RNA splicing)。内含子上游的剪接点称为 5′剪接点或左剪接点，下游方向的剪接点称为 3′剪接点或右剪接点。在有些内含子中，含有 4 个重复的保守序列，长度为 10～12 个碱基。4 个保守序列构成一种二级结构，在剪接中有重要作用，称为中部核心结构(central core structure)。含有中部核心结构的内含子称为第一类内含子(group I)，不含该结构的称为第二类内含子(group II)。这两类内含子的分布很广，特别是第一类内含子在细胞器基因和核基因中都有，了解比较详细的是纤毛原生动物四膜虫(*Tetrahymena*)的 rRNA 的内含子。第二类内含子如酿酒酵母线粒体细胞色素 c 氧化酶 a 亚基和 b 亚基的内含子 aI1、aI2、aI4 和 bI1。另一类是核基因 mRNA 前体的内含子，除含有第一类和第二类内含子外，还有一种具有 GA-AT 特征的边界序列。最后是 tRNA 基因的内含子，它们均位于反密码环。前三种内含子在剪接机制上有共同的特点，而 tRNA 内含子的剪接机制各有所异。不同的内含子与外显子的边界序列有一定的共同特征，如表 5-3 所示。

表 5-3　不同内含子的边界序列

内含子类型	5′剪接点	内含子	3′剪接点
酵母 tRNA 内含子	无保守序列	富含 A、U	无保守序列
第 I 类内含子(线粒体、核)	↓U		G↓
第 II 类内含子(线粒体)	▾GUGCG		YnAG↓
核 mRNA 内含子	C(A)AG↓GU(A)AGU		YnAG↓
酵母核 mRNA 内含子	▾GUAUGU		YnAG↓
眼虫叶绿体 mRNA 内含子	↓GUGU(C)G		YnAU(C)▾

内含子的去除及 RNA 的剪接基本有三种方式：在高等真核生物，核 RNA 内含子的去除由剪接装置(splicing apparatus)完成，在外显子和内含子交界处和内含子内部有可供识别的特异序列，剪接装置由多种蛋白质和核蛋白组成；有两类内含子的去除属自剪接，具有中部核心结构，形成特定的二级结构，RNA 具有催化剪接的作用；酵母 tRNA 的剪接需要蛋白质酶的参与，该酶与 tRNA 后加工过程中的酶有相似之处。除 tRNA 的剪接之外，其他 RNA 的剪接尽管参与反应的要素不同，但都属于转酯反应。

尽管已对 RNA 的剪接过程和机制有了一定的了解，仍有很多问题还不十分明确，如在核内剪接是否有特殊的定位，剪接是否与 RNA 其他的生物学过程有关系，剪接是否可造成不连续基因和连续基因在基因表达上有所区别，以及如何保证剪切位点选择的正确性等，都有待于进一步探索。

一、核基因 mRNA 的剪接

mRNA 的基因在低等真核生物中只有少数含有内含子，为不连续基因。而随着进化程度的增高，不连续基因的数目不断增加。在典型的哺乳动物基因中，约有 7～8 个内含子，分布在 16 kb 左右的范围内。外显子较短(100～200 bp)，内含子较长(1 kb)。初始转录产物称为 mRNA 前体，与基因是等长的，去除内含子后成熟的 mRNA 约为 2.2 kb。在细胞核中有一类核 RNA，十分不稳定，平均长度比 mRNA 要长，序列的复杂程度非常高，称为 hnRNA，其中包括了 mRNA 前体。hnRNA 以与蛋白质结合的状态存在，形成 hnRNP。体外研究表明，hnRNP 的形状为一个球体和一个与之相连的纤维状结构。头部为直径为 20 nm 的球形颗粒，沉降系数为 40 S。主要成分为 100～800 个核苷酸的 RNA 和多种蛋白质，蛋白质的分子量从 34 到 120 kDa。其中，6 种(A1、A2、B1、B2、C1 和 C3)蛋白质称为核心蛋白。每种蛋白质在一个球形结构中可有多个分子，蛋白质的种类在 20 种左右。其精确的结构及 RNA 以这种形式存在的作用还有待进一步探讨。

在 RNA 转录生成以后，在核内完成加帽和加尾的同时，内含子的去除即剪切过程也在进行。然后 RNA 转运到胞浆。

1. 核基因 mRNA 的剪接位点

核基因 mRNA 的剪接位点(splicing sites)是内含子与外显子的断裂点及相邻内含子的连接点。经过比较 mRNA 与其基因的序列，可以发现内含子和外显子的接头点有一些特点：一个内含子的两端没有广泛的序列同源性或互补性，但有形成二级结构以使剪接点相互靠近的可能；剪接点有一短的十分保守的序列：外显子 $A_{64}G_{73}-G_{100}T_{100}A_{62}A_{68}G_{83}T_{63}\cdots 12PyNC_{65}A_{100}G_{100}$－外显子。在内含子两端分别有两个非常保守的碱基，左剪接点为 GT，右剪接点为 AG。下面的数字表示该位置碱基所出现的百分数。内含子在剪接位点的这种特征又称为 GT-AG 规则。点突变研究证明，剪接位点 GT 或 AG 的突变可以阻止剪接的出现。左(5′)剪接点又可称为供体位点，右(3′)剪接点可称为受体位点。几乎所有的真核细胞核基因的内含子均遵循 GT-AG 规则，这也暗示着这类内含子存在着共同的剪接机制。线粒体、叶绿体和酵母 tRNA 的内含子不遵循这一规则。

典型的哺乳动物的 mRNA 有多个内含子。要将这些内含子全部去除并使外显子正确的连接起来，肯定存在着十分精确的作用机制。可能会有以下两种机制：RNA 本身有一种性质使内含子末端的某一点连接起来；所有的 5′剪接点具有相同的功能，所有的 3′剪接点没有结构上的区别，剪接遵循着某一规则，保证 5′剪接点正确的与 3′剪接点相连。

上已提及，核 RNA 的剪接位点及其周围并没有自身的互补序列，但在实验中，5′剪接点总是能正确的与 3′剪接点相连。如将 SV40 早期转录单位的第一个外显子连接到 β-球蛋白的第三个外显子后面，通过剪接仍能产生 SV40 外显子和 β-球蛋白外显子的杂合体。这说明：剪接位点具通用性，而内含子的其他序列(包括二级结构)不提供特定的信息，对剪接不产生影响；剪接装置没有组织特异性，RNA 只要含有正确的剪接点，在任何细胞中都可正确的剪接，但也有些例外的情况。按上述剪接的机制去推理，任一 5′剪接点可与任一 3′剪接点发生剪接，但实际上情况正好相反，剪接只发生在同一个内含子的

5′和 3′剪接点间，其机制尚不明了。

在 RNA 剪接过程中，内含子的去除是否有特定的顺序，也是需要了解的问题。RNA 印迹法(RNA blotting)可用于分析核 RNA 剪接过程中的中间体，从而确定内含子去除的顺序。通过对卵粘蛋白 mRNA 的印迹可得到 4 个明显的条带，说明剪接有特定的途径。如果内含子的去除是随机的，理论上应有 300 种以上的中间体出现，印迹不会出现上述结果。可以发现含不同内含子的中间产物，因此内含子的去除似没有特定的顺序，但也发现有一定的规律，当只去除一个内含子时，该内含子总是内含子 5 或 6；当去除两个时，包括内含子 5 或 6 的情况占多数，但另一个不甚确定。内含子 3 很少在剪接的前三步被去除。上述内含子去除的次序可以总结为 5/6，7/4，2，3/1。可以肯定的说还存在其他的方式，也有第 4 或 7 个内含子最后才去除的情况。通过上述实验可以看出，RNA 的构象影响剪接点的选择，在去除特定的内含子后，构象发生一定的变化，再选择新的剪接点。但内含子去除有不同的次序说明同一剪接阶段可有不同的构象。对于更长的基因，要分析构象对剪接的影响会更加困难。另外，剪接反应并不是按内含子在 RNA 前体上的顺序进行的。有一种模型认为剪接装置与 5′剪接点结合后，沿 RNA 前体分子向 3′剪接点移动，这样就保证了剪接的正确性。但也有一些现象该学说不能解释。

2. 核基因 mRNA 剪接的机制

体外剪接系统是研究剪接机制的重要手段，通过该系统可以研究 RNA 剪接的中间产物及某种成分在剪接中的作用。核提取物能完成 RNA 前体的剪接，说明剪接与转录过程无关。剪接与 RNA 的修饰状态也没有关系，没有帽子结构或 poly(A)尾的 RNA 都能完成剪接。

剪接可以分为三个阶段进行。第一阶段，内含子的 5′端切开，形成游离的左侧外显子和右侧的内含子-外显子分子。左侧的外显子呈线状，而右侧的内含子-外显子并不呈线状。在距内含子的 3′端约 30 个碱基处有一高度保守的 A，称为分支位点(branching site)，而内含子游离的 5′端以 5′-2′磷酸二酯键与 A 相连，形成一个套索(lariat)结构。第二个阶段，内含子的 3′剪接点被切断而以套索状释放，与此同时右侧外显子与左侧外显子连在一起。第三个阶段，内含子的套索被切开，形成线状并很快被降解。

完成剪接所需的条件只有三个序列，即 5′GT、3′AG 和分支点序列。经突变实验证明，内含子或外显子其他序列的缺失并不影响剪接。也说明剪接过程不需要内含子或外显子的特定构象。分支点周围的序列也有一定的保守性，在酵母保守性很强，其一致序列为 UACUAAC。其他高等真核生物保守程度差一些，但也有一定的规律，一定位置倾向于或是嘌呤或是嘧啶，其一致序列为 $Py_{80}NPy_{80}Py_{80}Pu_{75}APy_{95}$。分支点位于内含子 3′剪接点 18～40 个核苷酸处。在酵母中，分支点序列的突变可阻止剪接过程。高等真核生物分支点序列有一定的灵活性，当分支点缺失时，可以选择其他的相类似的序列来替代，并能产生正确的剪接产物，但替代位点总是尽可能的接近原来的分支点。因此，分支点的作用是选择距之最近的 3′剪接点作为与 5′剪接点相连的靶位点。

套索是通过内含子 5′的保守 G 与分支点 A 形成磷酸二酯键而形成的，这两个碱基在所有的真核生物高度保守。在酵母中，分支点为 UACUAAC 的倒数第二个 A，这和上述规律也是相统一的。在剪接过程中，没有酶的参与，但存在一种机制在反应前对上述保守

的序列进行检查。如在酵母中，UACUAAC发生突变会阻止任何剪接反应。所以在切断内含子5′端前，必须首先识别分支点。在有些5′剪接点的突变中，只能形成套索结构。如内含子的第一个碱基由G变为A，可切断左侧外显子并形成套索，但反应不能进一步进行。分支点序列的突变可阻止3′剪接点的断裂。另外一些突变只是简单的阻断剪接反应。

根据3′剪接点突变对剪接的影响，可将内含子分为两类，一类内含子3′剪接点的突变不影响5′剪接点的断裂及套索的形成，而另一类内含子则抑制上述反应。上述实验说明，在剪接反应前，有某种机制对剪接所需的序列元件进行检查，只有这些元件都存在的情况下剪接反应才会开始，而任一元件的严重缺失都会抑制剪接反应。如果每个元件在适当的位置上以可识别的形式存在，剪接反应才能起始，但反应会停止在突变元件处。

剪接过程中的反应为转酯反应(transesterification)，即酯键由一个位置转移到另一个位置(如图5-8所示)。首先由UACUAAC的A对5′剪接点发动亲水攻击，然后由已释放的外显子的3′端攻击3′端剪接点。在剪接反应前后，磷酸二酯键的数目并没改变，只是由原来的两个5′-3′磷酸二酯键(内含子与两个外显子间)变为5′-3′(两个外显子间)和5′-2′(内含子本身的套索结构)磷酸二酯键。

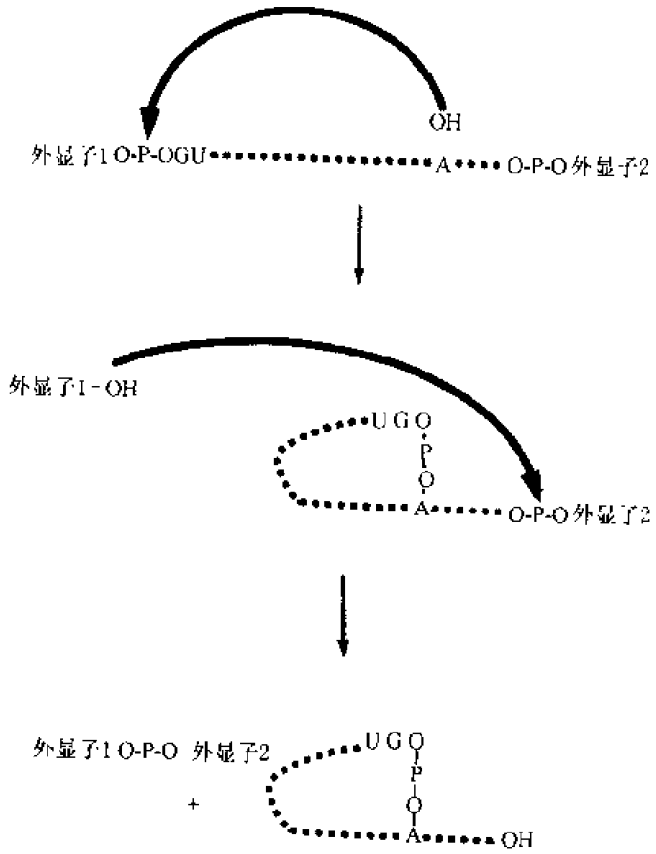


图5-8 核基因mRNA剪接过程中的转酯反应

剪接点的识别是剪接过程中的关键问题。可以设想有两种可能，一是通过酶来识别，一是通过RNA经碱基配对的方式形成二级结构，而后由酶识别。在后一种可能中，识别

剪接点的 RNA 可以来源于转录产物的本身(如线粒体的内含子),或来源于其他的 RNA 分子。在真核生物细胞的细胞核和细胞浆中,存在着很多的小分子 RNA,长度在 100～300 个碱基,在酵母可达 1 000 个碱基。它们的丰度有很大区别,含量最多的可达每个细胞 10^5-10^6 个分子,含量最小的用生物化学方法检测不到,只有用功能实验才能检测到。细胞核中的小分子 RNA 称为 snRNA,位于胞质中的称为 scRNA(small cytoplasmic RNA)。它们均以核蛋白颗粒的形式存在,分别称为 snRNP 和 scRNP,又俗称为 snurps 和 scyrps。一个 snRNP 颗粒一般含一个 RNA 分子和 10 个左右的蛋白质分子,其中有些蛋白质是某种 snRNP 颗粒所特有的,有些是几种所共有的。通过在酵母进行的突变实验,可以证明 snRNA 分子的重要功能,特别是那些参与剪接的 snRNA,这些 RNA 分子基因的突变是致死突变。有些参与剪接的 snRNA,在高等真核生物属较高丰度的,但在酵母却属较低丰度的。在参与剪接的 snRNP 分子中,RNA 含有与 3′或 5′剪接点或分支序列互补的序列,这种 snRNA 与 mRNA 前体或 snRNA 之间碱基配对在剪接中起重要作用。

各种参与剪接的成分形成一个剪接体系,称为剪接体(spliceosome),大小在 50～60 S,与核糖的大亚基相似。该体系由几种 snRNP 和大量的其他的蛋白质分子组成,这些蛋白质分子称为剪接因子,估计有 40 多种。snRNP 主要有 U1、U2、U5 和 U4/U6,U1、U2 和 U5 snRNP 分别只含有一个 snRNA 分子和约 20 种蛋白质,U4/U6 snRNP 含 U4 和 U6 两种 snRNA,同时含有几种蛋白质。所有的 snRNP 所含的蛋白质约有 40 余种,其中一部分直接参与剪接,另外一些可能提供特定的结构,或是起组装颗粒的作用,或是起联系颗粒的各种成分的作用。每种 snRNA 的核心由一组蛋白质构成,能被其抗血清识别,说明这些蛋白质存在保守序列。所有 snRNA 在动物、鸟类和昆虫有一定的保守性,且在酵母中都能找到相对应的 snRNA 分子,只不过酵母中长得多,但都含有相对保守的区域和特征。剪接点和分支点序列由剪接体识别,snRNA 和蛋白质都参与了识别,特别是 snRNA 之间及与 mRNA 间的碱基配对起重要作用。剪接体按一定顺序组装,已分离到一些组装的中间体。只有组装完整的剪接体才有功能。在剪接反应中,剪接体还会释放和添加某些成分。

人类的 U1 snRNP 含有 U1 snRNA 和 8 种蛋白质。U1 snRNA 有几个功能区域(如图 5-9 所示),Sm 结合位点(抗体结合位点)可与 snRNP 所共有的蛋白质相结合,由茎环结构所组成的几个功能区域与 U1 snRNP 特有的蛋白质结合。

真正参与剪接的区域位于 5′端的 11 个核苷酸,呈单链状态,含有一可与内含子 5′剪接点互补的序列,通常可形成 4～6 bp 的互补区域。在体外,完整的 U1 snRNP 颗粒能与 5′剪接点结合,但纯化的 U1 snRNA 没有结合 5′剪接点的能力,说明 snRNP 颗粒中的其他成分保证了这一结合。

经过对 U1 snRNA 参与腺病毒 12 SmRNA 前体剪接的突变实验,可以进一步明确二者间的相互作用。正常的 U1 snRNA5′端的序列为(5′)UACUUAC(3′),其中 3′的 AC 及 CU 可以和 12 SmRNA 前体的 5′剪接点附近序列(3′)GGAGUG(5′)的 UG 及 AG 配对:

U1 snRNA:3′CAUUCAU5′

12 S mRNA:5′GUGAGG3′

如果将 3′的 GG 变为 AU,即:

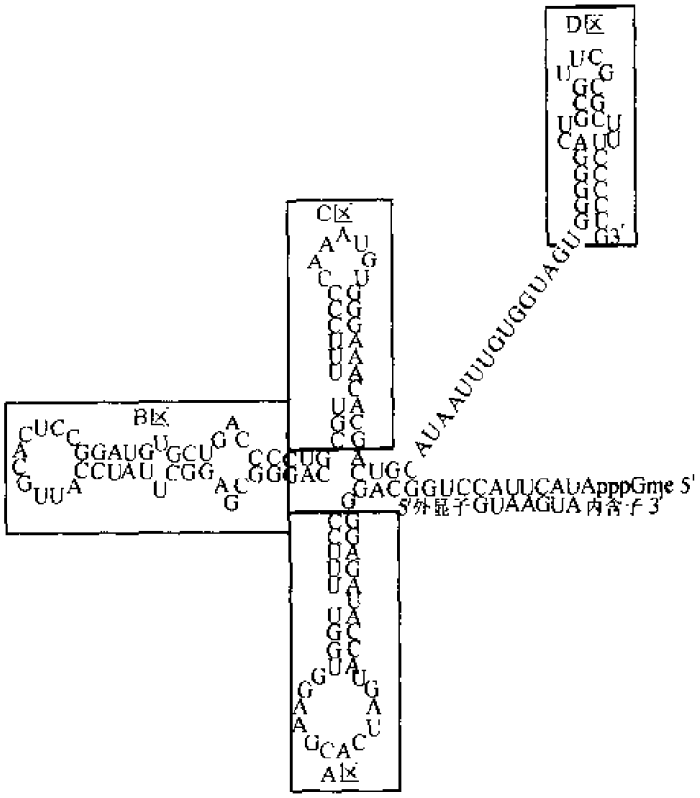


图 5-9　人类 U1 snRNA 的结构

U1 snRNA:3′CAUUCAU5′

12 S mRNA:5′GUGAAU3′

则剪接不能正常进行。如果再将 U1 snRNA 的倒数第三个 C 变为 U,即:

U1 snRNA:3′CAUUUAU5′

12 S mRNA:5′GUGAAU5′

剪接又能正常进行。在第一次突变中,没有改变二者间碱基配对数目,只是改变了碱基配对的位置,剪接就会因此而被抑制。第二次突变改变了 U1 snRNA 的一个碱基,使二者间的配对关系变为 4 个,但其中包括了正常状态下的三个配对碱基,尽管有一对碱基的种类发生了变化,但剪接仍能正常进行。这说明剪接的进行需要 U1 snRNA 与剪接点的碱基配对的关系,但并不仅仅依赖于配对的数目,更依赖于配对的位置。

U2 snRNA 含有能与分支点附近序列互补的顺序,位于其 5′端附近。用 U2 snRNA 的抗体进行免疫沉淀可分离到 U2 snRNA 与内含子的结合体。

核基因 mRNA 前体剪接的第一步是 E 复合物(早期前剪切复合物)的形成,该复合物含有 U1 snRNP 及剪接因子 U_2AF(又称 U_2 辅助因子)。E 复合物的形成确定了要剪接的 mRNA 前体。U1 snRNA 与内含子的 5′剪接点通过碱基配对结合。这一结合同时也需要 U1 snRNA 与分支点序列的结合,因为分支点序列的突变可阻止 U1 snRNA 与剪接点的结合。U1 snRNA 与分支点序列结合的机制尚不十分清楚,可能需 snRNP 颗粒中的其他成分或其他辅助因子参与。U1 snRNA 可能只起识别 5′剪接点的作用,在剪接体系

组装完成后，可以解离下来而不参与切断剪接点的过程。在上述结合完成后，U2 snRNA才能与分支点序列结合。U2 snRNA与E复合物的结合形成A复合物。U2 snRNA通过碱基配对来识别分支点并与之结合。U_2AF剪接因子与分支点下游的富含嘧啶的区域结合，U_2AF对于U1 snRNA和U2 snRNA均有辅助作用。在酵母中，U2 snRNA与分支点序列UACUAAC形成7个碱基对，这过程需剪接因子U_2AF，并需ATP供能。上已提及，分支点位于内含子的3′端，U1及U2 snRNA的结合，可使得两个剪接点在空间位置上相互接近，完成了剪接的起始阶段。在这一过程中，U1和U2 snRNA(RNP)起关键的作用。在二者失活的情况下，5′剪接点将不能被切开，套索结构也不能形成。U1 snRNA与5′剪接点结合后，U_2AF才能与分支点结合，并使得5′和3′剪接点互相靠近。剪接体中还有一组蛋白质成分，称为SR蛋白(SR protein)属于剪接因子和调解因子。目前已鉴定36种，含有Arg-Ser富含区，长度有所差异。与RNA结合的序列特异性差。可形成RNA底物的框架结构，将U_2AF与U1连接在一起。体外实验中，在没有U1的情况下，SR蛋白能将U_2AF与U2 snRNA结合。有些3′剪接点能与U_2AF和U2 snRNP有效的结合，需要有其他的序列与SR蛋白结合，来协助U_2AF与分支点的结合。这些序列称为剪接增强子(splicing enhancer)，常位于3′剪接点外显子的下游。

在内含子较长和剪接点较弱时，剪接体的形成可能还存在其他的途径。5′剪接点依然靠U1 snRNP识别，但3′剪接点的识别则是通过一个复合物来识别，该复合物包括了下一个外显子，该外显子的5′剪接点也由U1 snRNP结合，并经SR蛋白与U_2AF连接。在U2 snRNP结合形成A复合物后，会产生重排现象，使复合物与正确的5′剪接点结合。U5和U4/U6 snRNP三聚体与A复合物的结合形成B_1复合物，这时复合物中已包括了所有剪接所需的成分，因此可将B_1复合物看成是剪接体。接下来U1 snRNP释放，B_1复

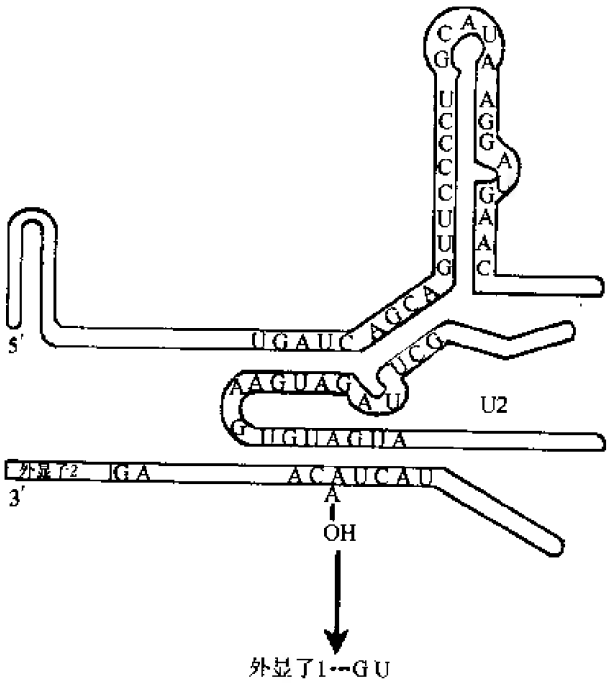


图5 10 U2与U6、U2与内含子的碱基配对及产生的活性位点

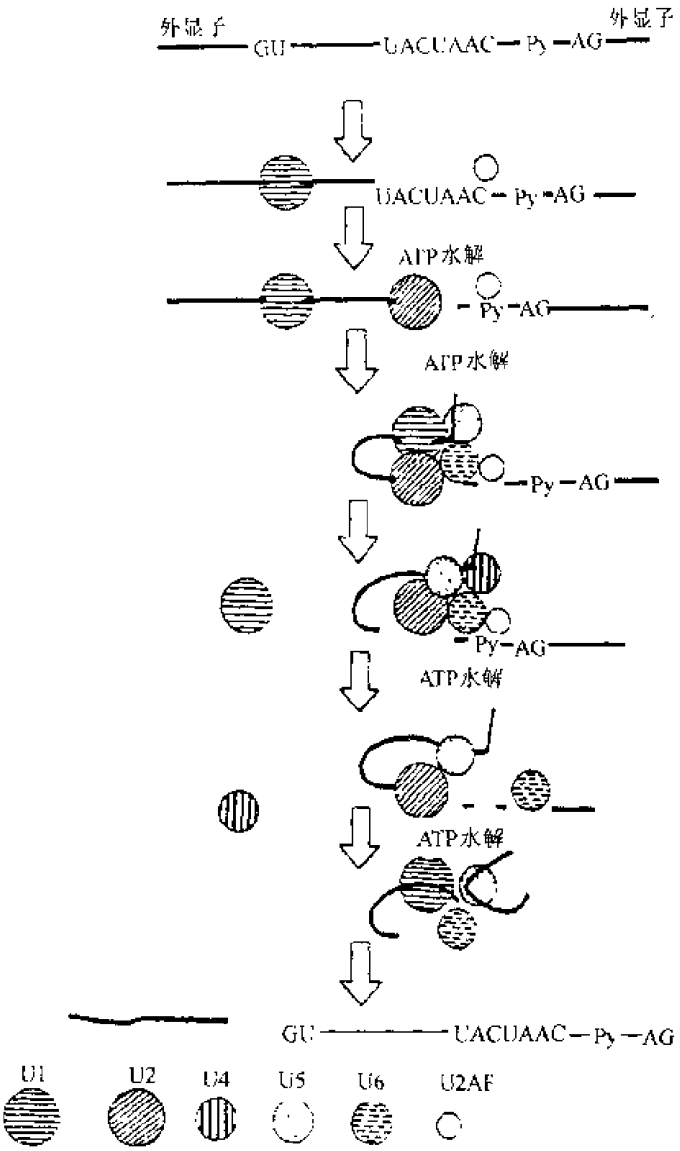


图 5-11　核基因 mRNA 的剪接

合物转变为 B_2 复合物。还不清楚 U1 snRNP 是如何脱离复合物的。U1 snRNP 的释放可引起剪接体中各种成分空间位置的变化，特别是 U6 snRNP。U5 snRNP 原来靠近 5′剪接点，U1 snRNP 释放后移到了内含子其他的序列上。

U4 snRNA 的释放会触发催化反应的进行。在此，U4 的功能可能是暂时封闭 U6 的功能。在 U4/U6 snRNP 中，U6 有 -26 个碱基的连续区域与 U4 的两段序列互补。U4 的解离可使 U6 的这段序列游离并呈另一种结构形式，其中有一部分与 U2 形成双链结构，另一部分自身形成发夹结构。由此可见，U4 的释放控制着剪接的进行。U6 snRNP 与 5′剪接点的保守顺序(GU)结合，有时还与 GU 下游的一段序列碱基配对。snRNA 之间及与 RNA 底物间存在着多种配对关系，如 U1(UCCAUUCA)与 5′剪接点(UGUAUG-

GA)、U2(UAGUA)与分支点(UACUAAC)、U6(AC)与5′剪接点(GU)及U5与两个外显子等。在snRNP之间和snRNP与底物RNA之间存在一些相近的序列。在环部结构中存在单链区。这些要素均是剪切所必需的。

剪接体中起催化转酯反应的成分尚未弄清楚,也不知道是蛋白质还是RNA在起作用,只是推测U2和U6形成的RNA结构可能是其中的成分之一,U2与U6及U2与内含子间的相互作用会产生具有催化活性的区域和位点(如图5-10所示)。套索结构的功能是定位3′剪接位点。在第一次转酯反应即5′剪接点断裂和套索结构形成后,立即进行第二步转酯反应,该反应需U5 snRNP与3′剪接点的结合。但目前尚不清楚U5 snRNP是否通过碱基配对来识别靶序列的,也没有发现二者间有互补序列。蛋白质成分可能参与了此过程,已鉴定到了一种蛋白质可与该位点结合,它可能是snRNP的某种成分。整个剪接过程如图5-11所示。

通过上面的叙述,我们可以清楚的看出,在剪接过程中,剪接体中的各种snRNA间及snRNA与底物间的碱基配对是至关重要的,这些作用可引起结构的变化以利于剪接的进行,使参与反应的基团处于合适的位置,并可能产生具催化作用的活性中心。而且,上述结构的变化是可逆的,如U6在反应后还要与U2分离,以便与U4结合从而参与下一轮的剪接。

由参与剪接的各种snRNA和蛋白质形成一个很大的复合物,称为剪接体,它是一个50~60 S的核蛋白体颗粒。在剪接体中,很难发现线形的5′外显子和套索结构,说明3′剪接点的断裂和外显子的连接是一个非常迅速的过程。从作用形式上看,剪接体中涉及了RNA和RNA之间、RNA和蛋白质之间及蛋白质之间的相互作用,这一点上和核糖体十分相似。电镜下观察到的剪接体为25 nm~50 nm,包括U1、U2、U4、U5和U6 snRNP。剪接体中各种snRNP不只是具有单一的功能。在snRNP分子中,存在着不同的功能区域,行使不同的功能。如U1 snRNP能刺激U2 snRNP的结合,又能与5′剪接点结合。U2 snRNP分子中除含有与分支点结合的区域外,还有与剪接体中其他成分相互作用的区域。

在酵母的突变研究中,发现了一些与剪接有关的基因,称为PRP(pre-RNA processing)基因。它们所编码的产物有的是RNA结合蛋白,有的属于依赖ATP的RNA螺旋酶家族。有些PRP编码蛋白是剪接体的成分,有些是独立起作用的因子。PRP16和PRP22都属于依赖ATP的RNA螺旋酶,PRP16参与第二次的转酯反应,PRP22参与成熟的mRNA分子从剪接体中的释放。在转酯反应中,并不需要能量,ATP水解释放的能量可能是用于剪接过程中的某些构象变化。

二、第一类内含子的剪接

第一类内含子的分布很广。它存在于低等真核生物的rRNA基因,如四膜虫(*Tetrahymena thermophila*)。真菌线粒体中也常能发现这类内含子。非常有趣的是在T4噬菌体的三个基因中也发现了第一类内含子,这一发现使内含子的存在扩展到了原核生物。此类内含子的特点一是剪接属于自剪接(self-splicing,autosplicing);二是形成明显的二级结构,共形成9个茎环结构。内含子的长度差异很大,保守序列距剪接位点很远。

1. 第一类内含子所形成的二级结构

所有的第一类内含子均有特征性的二级结构，有的甚至有三级结构。在其一级结构中有四个保守序列，依次为 5′- P-Q-R-S - 3′，其各自的一致序列为 P：AUGCUGGAAA；Q：AAUCAGCAGG；R：UCAGAGACUAA；S：AAGAUAUAGUCC。图 5 - 12 是四膜虫第一类内含子的二级结构示意图。

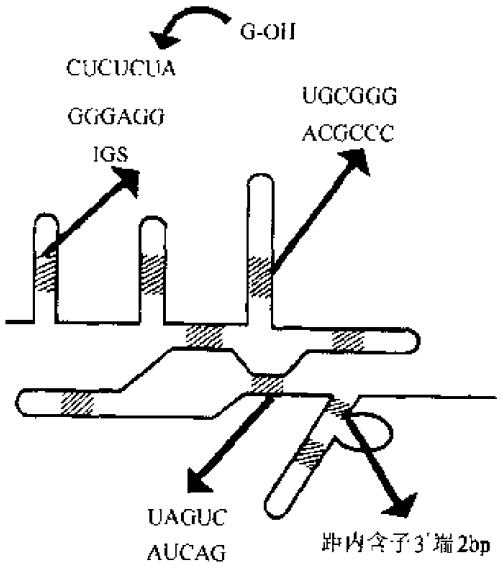


图 5 - 12　第一类内含子的二级结构

四膜虫第一类内含子的二级结构有 9 个区域(P1～9)。其中有两个由第一类内含子共同保守序列以碱基配对形成，P4 由 P 和 Q 间形成 6～7 个碱基的配对，P7 由 R 和 S 间形成 5 个碱基的配对。其他的茎环结构也是由两段互补序列形成，但这些序列没有发现有保守性，不同的内含子中序列各异。内含子中有一段序列可与外显子的某一序列互补，称为内部引导序列(internal guide sequence，IGS)，P1 即由二者形成。突变分析证明 P3、P4、P6 和 P7 属于核心结构，是能体现出催化活性的最小区域。

上述二级结构的形成对于剪接非常重要，有些碱基的配对可为剪接提供合适的构象。P1 中包括了左侧外显子的 3′端和内含子的 IGS。原来认为 IGS 可将剪接点拉在一起，现在看来这一反应并不是剪接所必需的。P7 和 P9 间的一个非常短的序列，只有 2 个碱基，可与内含子 3′端的 G 前面的 2 个碱基配对。

目前已建立了一个系统，将四膜虫的第一类内含子引入 β-半乳糖苷酶并成功的在大肠杆菌完成剪接。该系统的建立为研究内含子的核心序列在剪接中的作用奠定了基础。同时也说明剪接可在与四膜虫的内环境相差较大的大肠杆菌细胞中进行。在细菌体内的剪接过程，是否有细菌蛋白质及其他因素的参与还不清楚。

在应用上述系统对内含子的突变研究中，发现保守序列的突变可终止剪接反应，使突变链的互补链发生相应的突变而恢复链间的配对关系，可使剪接能重新顺利进行，成为前一种突变的回复突变。在线粒体的第一类内含子的研究中也得到了相同的结果。这说明

第一类内含子的剪接依赖于内含子内部所形成的二级结构，远离剪接点的序列可为自剪接的进行提供活性位点。

2. 第一类内含子剪接的机制

自剪接首先发现于四膜虫的 rRNA 的剪接。四膜虫的两种 rRNA 转录在一条 35 S 的产物中，小的 rRNA 位于前体的 5′侧，大的(26 S)rRNA 位于 3′侧。35 S 前体在体外就有自剪切的能力。在有些种类的四膜虫，26 S rRNA 的基因中只有一个内含子，该内含子切除后是 400 个碱基长的线状片段，而后变成环状分子。该反应只需要一价离子、二价离子和鸟苷酸辅助因子，GTP、GDP 和 GMP 均可，只要求鸟苷酸的 3′-OH。反应不需要外部能量的供给。用放射性同位素标记鸟苷酸参与剪接反应，可发现放射性出现在切除下来的线状内含子片段中，进一步研究确定 G 与内含子 5′端以磷酸二酯键相连。

这类内含子的剪接包括三次转酯反应，如图 5-13 所示。第一次转酯反应由游离的 G 发动，G 的 3′-OH 攻击内含子的 5′末端，产生 G-内含子和外显子游离的 3′-OH 末端。第二次转酯反应由上一个外显子的 3′-OH 攻击另一个外显子，并与之相连，同时释放线状的内含子。这两次转酯反应及二者间和连接反应同时进行，因为从反应体系中分离不到两个外显子的中间产物。游离的线状内含子发生第三次转酯反应而形成环状。

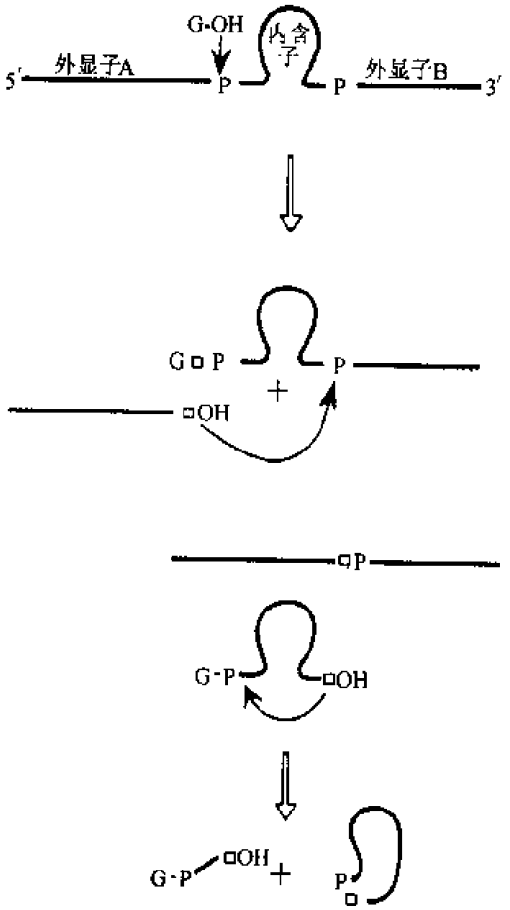


图 5-13 第一类内含子剪接过程中的三次转酯反应

自剪接反应的过程由转酯反应构成。转酯反应只是磷酸酯键的直接转移，没有水解反应出现，因此不需要外部的能量，也不需要供能物质如 ATP 或 GTP 的参与。内含子自身能形成特定的二级结构，从而产生适于剪接反应的空间构象和活性位点，在 G 的参与下完成剪接反应。这就是自剪接反应的最大特点，即剪接反应完全由内含子自身完成。当然，有时反应过程中也需要一些蛋白质分子，可能有维持 RNA 二级结构的作用。

自剪接过程中产生的游离内含子有下列性质：

1）线状分子具有自身环化的能力，3′末端的 G(G414)可对 5′端的两个碱基发动攻击并形成环化。多数情况下是 G414 与 A16 反应，称为基本环化反应(primary cyclization)少数情况与 U20 反应。环化后的分子呈环状，释放 5′端的 15(与 A16 反应)或 19(与 U20 反应)个碱基长的线状片段。

2）上述两种类型的环状分子在体外 G414 · A16 或 G14 - U20 可被水解，变为线状，这一过程称逆环化(reverse cyclization)。G414 · A16 产生的线状分子仍有形成 G414 · U20 环状分子的自身环化能力。环化形成的分子可被进一步水解，产生的线状分子称为 L - 19 RNA，该分子具有酶活性，可催化寡聚核苷酸的延伸。

3）自剪接释放的环形产物具有逆环化的能力，可与 UUU 反应，G414 - A16 的键断裂变为线状分子，并将 UUU 加到 5′端(A16)。该反应为分子内反应，表明该环形产物具有连接两个 RNA 分子的能力。

其他的第一类内含子的自剪接研究没有对四膜虫研究得详细，但也具有相似的特点。

第一类内含子在体外具有自剪接的能力，但在体内仍需要一些蛋白质的参与。在线粒体第一类内含子的剪接过程中，就需要一些基因产物(反式作用因子)的参与。例如，粗糙脉孢菌(*N. crassa*)的 *cyt18* 的突变可使其失去剪接线粒体第一类内含子的能力。该基因编码的是线粒体 tyr-tRNA 合成酶。这暗示内含子可能具有类似于 tRNA 的三级结构，可被 tRNA 合成酶识别。酶与 RNA 的结合可能直接参与剪接或是稳定 RNA 的构象。

三、第二类内含子的剪接

第二类内含子最早发现于线粒体，它的剪接位点与核基因 mRNA 的内含子有些类似，遵循 GT-AG 规则。在 3′剪接点上游 6～12 bp 的序列比较保守，为 PyPuPyPyTAPy，其中 A 绝对保守。在 A 的前后都有几个 3～5 个碱基的短序列，能与上游的部分序列形成二级结构，但 A 总能独立存在而不参与二级结构的形成。剪接机制与核基因 mRNA 也相似，形成套索状结构。剪接过程中发生两次转酯反应，第一次由上述保守的 A 攻击 G，第二次由 G 攻击 3′剪接点，完成转酯反应(如图 5 - 14 所示)。

分离到的第二类内含子在体外就能完成剪接反应，说明它的剪接过程不需要其他因素的参与，而 RNA 本身就具备完成剪接的能力，因此也属于自剪接。剪接过程不需要外部能量。第二类内含子可形成 6 个茎环结构即 6 个区(domain)，5 区和 6 区相隔 2 个碱基，上述保守的 A 就位于 6 区。A 的 2′- OH 发动第一次转酯反应，因此可将该区域视为 RNA 的有催化活性的区域。

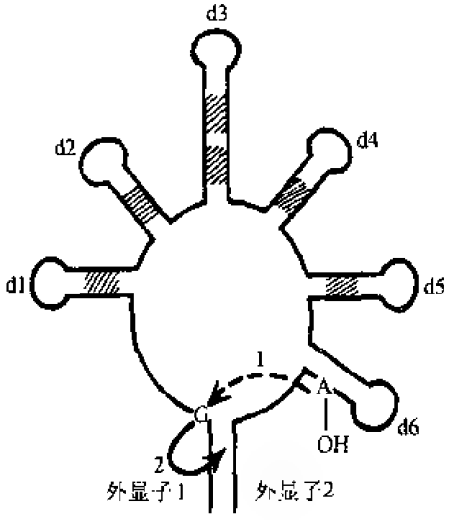


图 5-14　第二类内含子的转酯反应

与前已叙及的核基因 mRNA 的剪接过程相比，第二类内含子剪接最大特点是不需要其他成分的参与，RNA 分子本身能形成剪接所需的空间结构和催化活性区域。在核基因 mRNA 的剪接过程中，需要多种成分特别是 snRNA 与 RNA 前体共同作用才能形成剪接所需的空间结构，产生具催化功能的区域。这可能是生物进化的结果，将剪接功能由 RNA 本身转移到了其他的 RNA 分子乃至蛋白质分子。

四、酵母 tRNA 的剪接

上面讲述过的剪接反应均是依赖于一些短的保守序列，转酯反应与连接反应同时进行。在酵母 tRNA 的剪接过程中，采用了不同的机制，链的断裂与连接是两个独立的反应。

在 400 个左右的酵母 tRNA 基因中，有 40 个属割裂基因，均只含有一个内含子，位于反密码环 3′端，相距只有一个核苷酸，长度不一，在 14～16 bp 之间。功能相同的 tRNA 的内含子有着相似的序列，但携带同一氨基酸的 tRNA 的内含子的序列却大相径庭。在所有的内含子中，并没有发现可被剪切酶识别的同源序列，无论是在植物和两栖类，还是在哺乳动物都是这样。所有酵母的 tRNA 的内含子含有一段与反密码环互补的序列，序列互补的结果是使反密码环处形成了一个与成熟 tRNA 分子不同的构象（如图 5-15 所示），而其他的结构均和成熟的tRNA相同。

在酵母 tRNA 的剪接过程中，内含子的序列和大小对剪接无太大影响，很多突变并不影响正常剪接。但有些突变也影响剪接的进行。如突变发生在内含子环部的一个配对碱基和茎部的一个非配对碱基，将会影响剪接；另外一些改变内含子碱基配对状态的突变也会影响剪接。上述现象说明了 tRNA 一级结构对剪接的影响不大。研究表明，tRNA 的剪接主要依赖于其二级结构的特点，即主要识别其二级结构，如接受臂与 D 臂间的距离、

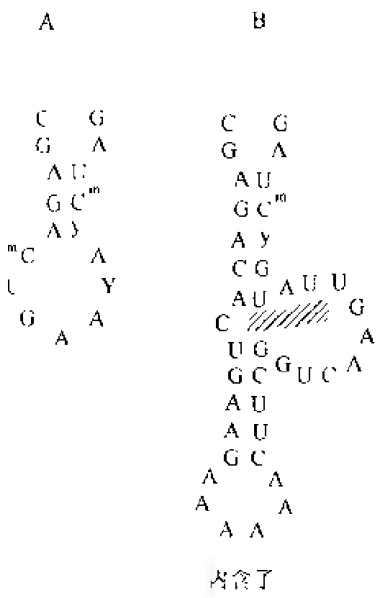


图 5-15 酵母 tRNA 前体的结构

TψC 臂的长度，特别是反密码子臂的长度。这和氨基酰 tRNA 合成酶识别相应的 tRNA 有些相似。

剪接主要分两步进行。第一步先由内切酶切断磷酸二酯键，该反应属非典型的核酸酶反应，不需要 ATP；第二步由 RNA 连接酶将断端连接形成熟的 tRNA 分子，该反应需要 ATP 的参与。内切酶在内含子的两端切断二酯键，形成一线状的内含子分子和两个 tRNA 半分子。tRNA 半分子有独特的末端，每个 5′端为羟(－OH)基，每个 3′端为 2′，3′环形磷酸基团。两个 tRNA 半分子经碱基配对形成一类似 tRNA 的结构。在加入 ATP 后，可启动第二步反应。环形磷酸基团在环化磷酸二酯酶的作用下打开，形成的产物具有 2′-磷酸基团和 3′-OH。5′-OH 必须被磷酸化为磷酸基团，才能在连接酶的作用和 3′-OH 连接形成磷酸二酯键。这时，内含子已被去除，但剪接点处的核苷酸仍有一个 2′磷酸基团，需在磷酸酶的作用下去除，最后形成熟的tRNA。

2′,3′环形磷酸基团在植物和哺乳动物的 tRNA 剪接中也出现。植物形成环形磷酸基团的反应与酵母基本相似，但动物的反应机理有很大不同。酵母的 tRNA 前体也可被卵细胞核提取物催化完成剪接过程，说明该提取物也具有识别和剪切酵母 tRNA 内含子的能力，剪接反应种属特异性较差。

研究 tRNA 剪接的方法主要是构建了一种温度敏感的酵母突变体，不能完成 tRNA 的剪接，tRNA 前体在细胞内堆积。提取这些 tRNA 的前体，在体外用细胞提取液对其进行剪接反应，从而了解各种成分在剪接中的作用及剪接的过程和机制。

（张永亮）

参 考 文 献

[1] Losick R et al. New York: Cold Spring Harbor Laboratory, 1976
[2] Doi R H et al. Microbil Rev, 1986. 50: 227－243
[3] Losick R et al. Ann Rev Genet, 1986. 20: 625－669
[4] Helmann J D et al. Ann Rev Biochem, 1988. 57: 839－872
[5] Rice G A et al. Proc Nat Acad Sci USA. 1991. 88: 4245－4249
[6] Haldenwang W G et al. Cell, 1981. 23: 615－624
[7] Grossman A D et al. Cell, 1984. 38: 383－390
[8] Haldenwang W G. Microbiol Rev, 1995. 59: 1－30
[9] Yong R A. Ann Rev Biochem, 1991. 60: 189－715
[10] Sakonju S et al. Cell, 1980, 19: 3－15
[11] Sakonju S et al. Cell, 1980, 19: 27－35

[12] Kassavatis G A. Cell, 1990. 60:235~245
[13] Buratowski S. Cell, 1989. 56:549~561
[14] Zenke M et al. EMBO J, 1989. 5:387~397
[15] Mueller-Storm H P. Cell, 1989. 58:767~777
[16] Stephen T Smale. Biochimica et Biophysica Acta, 1997. 1351:73~88
[17] Steve Busby. Cell, 1994. 79:743~746
[18] Adhya S et al. Ann Rev Biochem, 1978. 47:967~996
[19] Platt T. Ann Rev Biochem, 1986. 55:339~372
[20] Roberts T W. Nature, 1969. 224:1168~1174
[21] Greenblatt J et al. Cell, 1981. 24:421~428
[22] Yager R D et al. Neidhardt, American Society for Microbiology, Washington DC, 1987. 1241~1275
[23] Das A. Ann Rev Biochem, 1993. 62:893~930
[24] Geiduschek E P et al. Ann Rev Biochem, 1988. 57:873~914
[25] Zawel L et al. Prog Nucleic Acids Res Mol Biol, 1993. 44:67~108
[26] Burley S K et al. Ann Rev Biochem, 1996. 65:769~800
[27] Friedberg E. Ann Rev Biochem, 1996. 65:14~42
[28] Kim Y et al. Nature, 1993. 365:512~520
[29] Goodrich J A et al. Cell, 1993. 77:145~156
[30] Martinez E et al. EMBO J, 1994. 13:3115~3126
[31] Nikolov D B et al. Nature, 1995. 377:199~228
[32] Verrijzer L P et al. Cell, 1995. 81:1115~1125
[33] Leigh Zawel. Ann Rev Biochem, 1995. 64:533~561
[34] Stanley N C. Cell, 1995. 80:829~832
[35] Bannerjee A K. Microbiol Rev, 1980. 44:175~205
[36] Jackson R J et al. Cell, 1990. 62:15~24
[37] Ross J. Microbiol Rev, 1995. 59:423~450
[38] Gegenheimer P et al. Microbiol Rev, 1981. 45:502~541
[39] Kole R et al. in the enzymes, ed. Boyer P D. 1982. 15:570~483
[40] Sharp P A. Science, 1987. 235:766~771
[41] Guthrie C et al. Ann Rev Genet, 1988. 22:387~419
[42] Grthrie C. Science, 1991. 253:157~163
[43] Nilson T. Ann Rev Microbiol, 1993. 47:413~440
[44] Weiner A. Cell, 1993. 72:161~164
[45] Sharp P A. Cell, 1994. 77:805~815
[46] Kramer A. Ann Rev Biochem, 1996. 65:367~410
[47] Zhung Y et al. Cell, 1980. 46:827~835
[48] Zhang Y et al. Proc Nat Acad Sci USA, 1989. 86:2752 -2756
[49] Reed R et al. Cell, 1988. 53:949~961
[50] Sontheimer E J et al. Science, 1993. 262:1989~1996
[51] Hannon G J et al. Cell, 1980. 61:1247~1255
[52] Bruzik J P et al. Proc Nat Acad Sci USA, 1995. 92:7056~7059
[53] Cech T R. Ing Rev Cytol, 1985. 93:3~22
[54] Cech T R. Science, 1987. 236:1532~1539
[55] Cech T R. Ann Rev Biochem, 1990. 59:543~568
[56] Lambowitz A M et al. Ann Rev Biochem, 1993. 62:587~622

[57] Michel F et al. Ann Rev Biochem, 1995. 64:435～461
[58] Gert J A. Biochimica et Biophysica Acta, 1996. 1307:39～54
[59] Hiten D M et al. Ann Rev Genet, 1994. 28:1～26

第六章　蛋白质生物合成

蛋白质生物合成又称翻译(translation),是指由 RNA 参与的蛋白质生物合成的过程,它将核酸序列转变为氨基酸序列。参与翻译的 RNA 分子有 tRNA、rRNA 和 mRNA。tRNA 的功能是转运氨基酸,rRNA 与多种蛋白质组成核糖体做为翻译进行的场所,mRNA做为翻译的模板。经过三种 RNA 及多种蛋白质的相互作用,使 mRNA 由 DNA 而来的遗传信息正确的传递到蛋白质。

第一节　遗传密码[1~6]

一、遗传密码

遗传密码(genetic code)是联系核酸的碱基序列和蛋白质的氨基酸序列的途径。由三个碱基代表一种氨基酸,称为密码子(codon)。4 种碱基可以组合成 64 种密码子,而体内只有 20 种氨基酸,故存在多个密码子代表一种氨基酸的情况。

密码子的破译是分子生物学发展史上的一件大事。在此项伟大的工作中,主要运用了两项技术,一项是1961 年由 Nirenberg 报道的人工合成多聚核苷酸体外翻译技术;另一项是 Nirenberg 和 Leder 发明的核糖体结合技术。这两项技术的出现,为密码子的破译奠定了基础。随后人们经过很多精彩的实验,只用了几年的时间,就破译了全部的密码子,同时也确立了碱基序列与氨基酸序列间的关系。整个密码子字典如表 6-1 所示。

表 6-1　遗传密码子字典

第一位置(5′端)	第二位置(中间碱基)				第三位置(3′端)
	U	C	A	G	
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	终止	终止	A
	Leu	Ser	终止	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met*	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val*	Ala	Glu	Gly	G

* 为同时编码起始氨基酸

由密码子字典可以看出，除甲硫氨酸和色氨酸外，其他的氨基酸均有两种以上的密码子。多种密码子编码一种氨基酸的现象称为简并(degeneracy)，代表一种氨基酸的密码子称为同义密码子(synonyms)。氨基酸的密码子的数目和它在蛋白质中出现的频率间并没有明显的正相关性。

密码子表中有8个密码子家族(codon family)，每个家族有4种密码子，编码一种氨基酸，这时密码子的第三位碱基对密码子的性质没有影响。只有三个密码子的第三位碱基具有意义，即AUG、UGG和UGA。因此，第三位的C和U在任何密码子中均没有特殊的意义，而A在氨基酸的密码子中也是如此。一种氨基酸的几个密码子的差异通常在第三个碱基，说明了密码子的第三个碱基(third-base degeneracy)有一定的灵活性。这种灵活性和密码子的简并可使有些突变只引起核酸序列的改变而不引起蛋白质中氨基酸序列的改变，从而把突变对生物体造成的影响降低到最小。

在所有的生物体中，密码子字典几乎是通用的，这已被体内和体外翻译实验所证明。密码子的通用表明密码子可能在生命进化的早期就建立起来了。也有极少数例外的情况。在山羊枝原体(*Mycoplasma capricolum*)，UGA不是终止密码子，而代表色氨酸，而且使用频率比UGG高得多。共有两种Trp-tRNA，一种反密码子为ACU，识别UGA；一种为ACC，识别UGG。在有些种类的纤毛虫，UAA和UAG代表的是谷胺酰胺而不是终止密码子，在一种纤毛虫中，谷氨酸有三种tRNA，一种识别CAA和CAG，另一种识别UAA和UGA，最后一种识别UAG。在四膜虫中也有这种情况。在另一种纤毛虫(*Euplotes octacarinatus*)，UGA代表胱氨酸，UAG没有发现，只有UAA做为终止密码子。这些例外只是涉及了终止密码子。在假丝酵母(*Candina*)，GUG是丝氨酸而不是亮氨酸的密码子，这是目前发现的唯一的编码密码子的改变。线粒体的遗传密码存在较多的例外情况，其中最常见的一种变化是UGA由终止密码子变为色氨酸的密码子(植物线粒体除外)，这可能是在进化上最早期的变化。有些变化使得一些具有不同含义的密码子代表一个氨基酸，如UGG和UGA(终止密码子)均代表氨基酸，AUG和AUA(异亮氨酸)均代表甲硫氨酸。这使得线粒体的密码子更加复合密码子中两个碱基特异性强的特点。除植物外，酵母、脊椎和非脊椎动物均是如此。其他的变化发生在特定的一些物种。

密码子决定氨基酸是通过mRNA与tRNA相互作用实现的，这种作用实质上是碱基间的配对。游离的三聚体间的配对很不稳定。核糖体的A位有稳定作用，从而稳定了遗传信息的传递。

二、密码子与反密码子的相互作用

密码子与反密码子(anticodon)间的正确识别是遗传信息正确传递的保证。在生物体中，tRNA的种类要少于编码氨基酸的密码子(61种)，要识别这些密码子，必然存在一种反密码子识别多种密码子的现象。一般来讲，密码子的前两位碱基在和反密码子配对时遵循Watson-Crick原则(A与U、G与C)，而第三位碱基则有一定的灵活性，这就是摆动假说(wobble hypothesis)。tRNA反密码环的构象特征是这种灵活性的基础。碱基间的非正常配对如表6-2所示。

表 6-2 密码子与反密码子第三位碱基的摆动配对

反密码子的第一个碱基	可识别的密码子的第三个碱基	反密码子的第一个碱基	可识别的密码子的第三个碱基
U	A或G	C	G
A	U	G	G或U

摆动配对中G可与U配对，结果使得第三位的A对密码子的含义不产生任何影响。第三位的G或U仍对密码子的含义有影响，但只有UGG和AUG两个密码子属于这种情况。

tRNA分子中含有较多的修饰碱基。位于反密码子中的或其他位置的修饰碱基均会对密码子和反密码子的作用产生影响。实际上，有些正常的碱基配对并不出现在密码子与反密码子之间，如U和A基本不出现在反密码子的第一位。即使U出现在第一位，也会被修饰而改变配对性质；A也偶尔出现在第一位，但将被修饰成I。反密码子第一位为肌醇的几率较高，能与U、C或A配对，这对子异亮氨酸的密码子更为重要。异亮氨酸的密码子为AUA，第三位为A。按正常的碱基配对，需要反密码子第一位为U，但这样的密码子又同时能识别甲硫氨酸的密码子AUG，会造成翻译的错误。因此，AUA必须由AUU及AUC的反密码子识别，该反密码子第一位为I。有时修饰作用会使碱基与其他碱基的配对有一定的选择性。如反密码子第一位的5-氧乙酸尿嘧啶和5-甲氧尿嘧啶易子识别A和G，对U识别效率较低。G的修饰不改变它的配对性质，但会对U的结合更加容易。

在不同生物体内，存在多种方式使少数的tRNA识别61种编码氨基酸的密码子，不同的生物所采用的途径也不同。在有些生物，缺乏某种特定的碱基修饰方式，也就缺乏某种修饰碱基的配对方式。也就是说，不同生物中可有自身独特的反密码子识别密码子的方式。有些密码子可由两种以上的tRNA识别，但不同的tRNA识别的效率不同。

多种反密码子与密码子的识别都和摆动假说相一致，但也发现了一些不相符的例子。反密码子周围的碱基和反密码环的空间结构对识别密码子有一定的影响。在有些突变体中，tRNA一个位置碱基的变化会改变其识别的密码子。细菌中fMet-tRNAf对起始密码子AUG和GUG的识别，也不符合摆动假说。

按摆动假说推算，要识别61种密码子至少需要31种tRNA，而哺乳动物线粒体中只有22种tRNA，所以必定存在着特殊机制。如果一种tRNA可以识别一个密码子家族中的4个密码子，最少需要23种tRNA，而线粒体中AGG和AGA变为终止密码子，又可减少1种tRNA，正好为22种。识别8个密码子家族的tRNA在反密码子的第一位为U。其他的密码子可分为两类，一类以嘧啶结尾，由反密码子的G识别；一类以嘌呤结尾，由反密码子的U识别。这均符合摆动假说。UGG变为UGA的同义密码子(色氨酸)，AUA变为AUG的同义密码子(甲硫氨酸)。这种变化使非密码子家族的密码子变为14对，由14个tRNA识别。现已鉴定了22种编码tRNA的基因。

线粒体中另外一个十分特殊的情况是mRNA的结构也影响密码子的识别。大肠杆菌中有二个编码富马酸脱氢酶同工酶的基因，其mRNA中的一个UGA由硒代半胱氨酰tRNA(seleno-cys-tRNA)识别，而AUG下游发夹结构的存在是这种识别所必须的。真核

生物中也发现了类似的情况。含硒代半胱氨酸的蛋白质往往参与催化氧化还原反应，硒代半胱氨酸属于活性位点。4 个基因的突变可引起硒蛋白的合成障碍。*Sel C* 编码 tRNA，反密码子为 UCA，携带丝氨酸。Sel A 和 Sel D 负责将丝氨酸修饰成硒代半胱氨酸。*Sel B* 编码一种 GTP 结合蛋白，协助硒代半胱氨酸 tRNA 进入 A 位，其作用类似于 EF-Tu。氨基酸序列与 EF-Tu 和 IF-2 有同源性。

第二节　tRNA 的功能[7~17]

tRNA 在翻译中起转运氨基酸的作用，将氨基酸经过密码子与反密码子间的相互识别定位在肽链中。从这一点上看，tRNA 是起具体翻译作用的分子。所有的 tRNA 分子都有相同的二级（三叶草形）和三级（倒 L 形）结构。这种结构上的一致是其功能所必需的，因为不同的tRNA分子有相同的功能特征，如均能和核糖体的 A 位和 P 位结合；能被翻译因子（EF-Tu 或 eEF1）识别。起始 tRNA 由起始因子识别，它在结构上有自己的特点。

对于 tRNA 和蛋白质之间的作用，我们还没有一个系统化的认识。虽然已经得到了所有 tRNA 或是一类 tRNA 分子中的一些共同的短核苷酸序列，但还没能将这些保守的一级结构与蛋白质的识别联系起来。已确定了 tRNA 的某些区域参与了与氨酰 tRNA 合成酶的相互作用，但是具体的接触位点在不同的 tRNA 分子中有所区别。

一、氨酰 tRNA 合成酶

氨酰 tRNA 合成酶催化氨基酸与 tRNA 间的反应。携带一种氨基酸的 tRNA 可有多种，称为同工 tRNA(iso-accepting tRNA)。可将 tRNA 分为 20 个同工 tRNA 组。一组同工tRNA在特定的合成酶催化下与相应的氨基酸结合。同工 tRNA 在结构上有不同的特征，以供合成酶识别，但还没能确定这种结构特征。从 tRNA 一级结构来看，有一些同源性非常高的短核苷酸序列，但没能找到同源序列与不同的氨酰 tRNA 合成酶间有相关性。从高级结构来看，所有的 tRNA 又非常相似。

氨酰 tRNA 合成酶有三个结合位点，为氨基酸结合位点、ATP 结合位点和 tRNA 结合位点。它催化的反应可分为两步，第一步氨基酸和 ATP 形成腺苷酸化氨基酰，释放焦磷酸；第二步活化型的氨基酸被转移至 tRNA，释放 AMP。

氨酰 tRNA 合成酶所催化的反应有特异性，即能选择相应的 tRNA 和氨基酸。一种合成酶只能识别一种氨基酸和该氨基酸的几种同工 tRNA。同工 tRNA 间存在结构上的差异，反密码子也可不同。被一种合成酶识别的 tRNA 称为关连 tRNA(cognate tRNA)。

生物体有 20 种氨酰 tRNA 合成酶，分子量相差很大，在 40 kDa～100 kDa 之间，有单聚体、二聚体和四聚体。不同酶分子间几乎没有同源性，可能起源于进化的早期。识别 tRNA 的活性位点只占整个分子的极小部分。20 种合成酶按其 tRNA 结合区域的结构可分为两类，每一类各 10 种。合成酶分子中有 4 种活性区域，即催化区域、tRNA 接受臂螺旋结合区域、tRNA 反密码子结合区域和聚合区域。催化区域包括 ATP 和氨基酸结合位点，接受臂螺旋结合区域介入至催化区域中，这样使 tRNA 的 3′末端接近催化位点。在能

形成多聚体(第一类合成酶)的合成酶分子中有聚合区域,其功能是使酶分子聚合,第二类合成酶没有聚合方式。

第一类合成酶有一 N 端催化区域,含有两个短的、部分保守的氨基酸序列,有时称其为特征序列(signature sequence)。催化区域呈核酸结合折叠形式,由 β 折叠和 α 螺旋组成交替平行结构。ATP 结合位点包含了部分特征序列。介入到催化区域的 tRNA 接受臂螺旋结合区域在第一类合成酶中差异很大。C 端区域包括 tRNA 反密码子结合区域和聚合区域,不同第一类合成酶中有很大区别。第二类合成酶的催化区域中有三个同源性较高的序列,活性位点有一个大的反向平行 β 折叠结构,周围绕着 α 螺旋。tRNA 接受臂螺旋结合区域介入到催化区域间,不同酶分子该区域的结构不同。反密码子结合区域偏向 N 端,聚合区域的定位变化很大。不同种类合成酶分子间没有明显的相似关系,这是令人费解的。

tRNA 沿着 L 形的一面与合成酶结合。所有的 tRNA 均是以其两个端点与合成酶结合,大部分的序列并不被合成酶识别。第一类和第二类合成酶与 tRNA 的相互作用有所区别。经晶体结构分析,两类酶与 tRNA 接触的方位正好相反,所形成的复合物互成镜像结构。Gln-tRNA合成酶(属第一类合成酶)与 tRNA 的 D 环接触,同时在 端识别接受茎的小沟,在另一端识别反密码环(如图 6-1 所示)。复合物的形成会引起 tRNA 分子中两个重要部位空间结构的改变(如图中的虚线为原先的结构,实线为改变后的结构。)。一是反密码环的 U35 和 U36 被拉向合成酶;一是接受茎的末端严重扭曲,致使 U1 与 A72 的配对关系被破坏,茎的单链伸入合成酶分子表面形成的深口袋中,该口袋中也含有 ATP 的结合位点。在以前的研究中也已经发现,U35 及 G73 的突变和 U1 与 A72 的配对影响酶对 tRNA 的识别。tRNA 与合成酶通过氢键结合。

在 Asp-tRNA 合成酶(属第二类酶),氨基酰 tRNA 从相反的方向接触合成酶,识别可变环和接受茎的大沟(如图 6-2 所示)。接受茎的螺旋构象没有发生变化。ATP 可能结合于末端 A 的附近。在另一端,反密码环与酶分子结合非常紧密,并引起反密码环构象的变化。

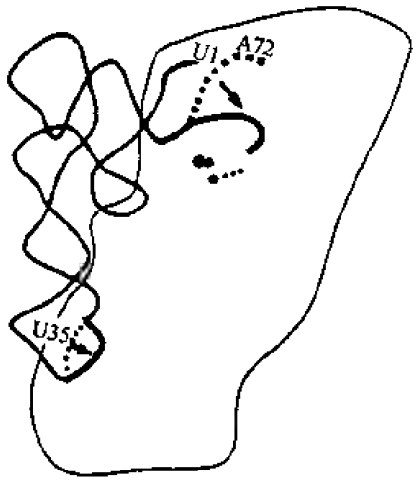


图 6-1 第一类合成酶与 tRNA 的结合

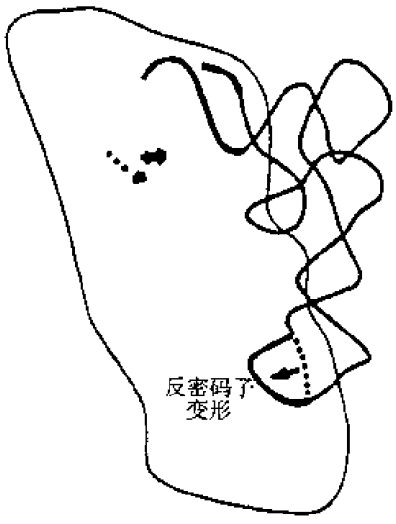


图 6-2 第二类合成酶与 tRNA 的结合

同工 tRNA 必须由一种氨酰 tRNA 合成酶所识别，因此它们应该具有某些共同的特征，并能与其他的 tRNA 区分开。人们一直在寻找同一种合成酶的 tRNA 底物间的结构相似性，以及通过结构的改变而改变 tRNA 携带氨基酸的性质，但没能找到一定的规律性。合成酶对不同 tRNA 识别的结构基础不同，也不单独依赖 tRNA 的一级或二级结构。在反密码环和接受茎的突变可改变合成酶的识别，但反密码子的突变并不影响识别。表 6－3 总结了一些 tRNA 被酶识别的结构特征。tRNA 分子中的关键位置很少，在 1－5 个之间。反密码环中经常有一些关键的核苷酸，通常反密码子中至少有一个被识别，有时三个碱基均很重要。接受茎的最后三个配对碱基中的一对被识别，如丙氨酸的 tRNA，只有接受茎的一对碱基受合成酶识别。在接受茎和 CCA 末端之间，同工 tRNA 常有些保守碱基，称为鉴别碱基(discriminator base)。但这些特征不足以使 20 组同工 tRNA 区分开。这更进一步说明了合成酶对不同 tRNA 的识别各自遵循着自己的规律。

表 6－3　一些 tRNA 被合成酶识别的特征碱基

tRNA	改变氨基酰 tRNA 合成酶识别的突变
I 类酶	
缬氨酸	反密码子的三个碱基的改变
半胱氨酸	反密码子的三个碱基的改变
异亮氨酸	反密码子环中 C34 的修饰
谷胺酰胺	反密码子环中 U35，接受茎 U1－A72 和 G73
II 类酶	
苯丙氨酸	反密码子的三个碱基；D 环的 G20；末端的 A73
丝氨酸	接受茎中 G1－C71；G2－C71；A3－U70；D 茎的 C11－G24
丙氨酸	接受茎中 G3－U70

几种合成酶可识别由接受臂和 TφC 臂组成的小螺旋结构，并催化其与氨基酸的结合。第二类合成酶对这种底物的氨酰基化的活性较高。对于这类 tRNA，特异性不依赖于接受茎，但不同的 tRNA 分子间小螺旋的结构差异很大。有些情况下，反密码子的突变会影响第二类合成酶对 tRNA 的识别。因此，在氨基酰－tRNA 合成酶与 tRNA 识别的过程中，涉及到了多种因素，且不同的氨基酰 tRNA 合成酶识别 tRNA 的部位是不同的。

二、抑制 tRNA

tRNA 突变体的分离，是研究反密码子与密码子相互作用的重要手段，也是研究 tRNA 分子不同部位对反密码子相互作用影响的有效途径。有些 tRNA 的突变，特别是反密码子的突变，可以抑制基因突变效应的产生，这类 tRNA 称为抑制型 tRNA(suppressor tRNA)。抑制型 tRNA 发挥作用的机理为：mRNA 分子上密码子的突变可造成其所编码的蛋白质的结构变化，导致蛋白质功能的丧失；tRNA 反密码子的突变使其可识别突变的密码子，蛋白质的结构不发生变化。后一种突变成为前一种突变的抑制突变。在野生型细胞，无义突变指密码子变为终止密码子而使翻译提前终止。抑制突变产生一种 tRNA

分子，可识别终止密码子，使翻译通过了无义突变而继续进行。但如果抑制型 tRNA 所携带的氨基酸与野生型不同，会影响蛋白质的功能。无义抑制突变可分为三类，分别识别三种终止密码子 UAA、UAG 和 UGA。识别 UAA 的为赭石型；识别 UAG 的为琥珀型；识别 UGA 的为乳石型。一些已明确的无义抑制突变如表 6-4 所示。

表 6-4　无义抑制 tRNA 的例子

基因定位	tRNA	野生型		抑制突变型	
		密码子	反密码子	反密码子	密码子
Sup D	Ser	UCG	CGA	CUA	UAG
Sup E	Gln	CAG	CUG	CUA	UAG
Sup F	Tyr	UAC UAU	GUA	CUA	UAG
Sup C	Tyr	UAC UAU	GUA	UUA	UAA UAG
Sup G	Lys	AA AAG	UUU	UUA	UAA UAG
Sup U	Trp	UGG	CCA	UCA	UGA UGG

最易鉴定的是琥珀型抑制 tRNA。在大肠杆菌中，至少有 6 种 tRNA 可识别 UAG。这些 tRNA 的反密码子为 AUC，由野生型改变反密码子中的任一个碱基而来。携带的氨基酸为丝氨酸、谷氨酰胺或酪氨酸，与野生型的 tRNA 相同。赭石型 tRNA 也是由反密码子的突变而来（如 supC 和 supG），可识别 UAA 和 UAG，在突变的密码子位置上插入酪氨酸或赖氨酸。Trp-tRNA 是识别 UGA 的抑制型 tRNA，野生型的密码子为 ACC，只识别 UGG。突变后为 ACU，即可识别 UGG，又可识别 UGA。Trp-tRNA 的另外一种乳石型突变反密码子没有发生变化，而是 G24 变为 A，使 D 茎的 GU 对变为 AU 对，增加了双螺旋的稳定性。除 UGG 外，该 tRNA 能识别 UGA，C 与 A 配对又不符合摆动原则，可能是 D 茎的变化引起了反密码环构象的变化，改变了反密码子的配对性质。

真核生物也发现了类似的现象。如牛肝中有一种 Ser-tRNA，反密码子为 mCCA。按摆动原则，可以识别色氨酸的密码子 UGG，但实际上却可识别 UGA。因此，可能是 UGA 的抑制型 tRNA。从上面的论述可以看出，抑制型 tRNA 和终止密码子间的识别并不能只是依赖于反密码子和密码子的序列，而是受 tRNA 分子内其他因素的影响。

上述的抑制型 tRNA 突变是由编码 tRNA 基因的点突变引起的。从理论上讲，对反密码子碱基的修饰的变化也会引起类似的突变，如编码修饰 tRNA 的酶的基因突变。

另外，抑制型 tRNA 同样也可抑制错义突变。如某一密码子发生突变，能识别突变密码子且携带野生密码子编码的氨基酸的 tRNA，就能抑制这种突变的效应。有两种方式可以形成抑制型 tRNA，一是 tRNA 反密码子发生突变，一是 tRNA 其他的结构变化或是氨酰 tRNA 合成酶的突变而改变了其荷载氨基酸的变化。后一种情况较少见。在大肠杆菌和其他细菌中都发现了错义抑制突变，对真核生物这方面的情况了解较少。

和正常的反密码子与密码子间的相互识别相比，抑制型 tRNA 与反密码子间的识别有不同之处。正常情况下，密码子只有一个含义；在抑制突变体中，密码子可被抑制型 tRNA 识别，也可被正常的 tRNA 识别。因此，无义突变抑制型 tRNA 必须和释放因子竞争与终止密码子的结合，错义突变抑制型 tRNA 要和正常 tRNA 竞争与密码子的结合。

竞争的程度会影响对突变抑制的效应。所以，抑制型 tRNA 的效应不只是依赖反密码子与密码子的亲合力，也受本身在细胞内的浓度和其他因素的影响。

无义突变抑制型 tRNA 的效率随密码子在 mRNA 中的位置变化而有很大不同，密码子的上下文序列会影响抑制效应。现还不明确密码子相邻的碱基对其和反密码子作用的影响，但相邻碱基的变化可使抑制型 tRNA 的抑制频率改变一个数量级以上，3′端碱基影响比 5′端大得多。

无义突变抑制型 tRNA 在抑制无义突变的同时，也会使 mRNA 分子中正常的终止密码子编码氨基酸，从而产生通读，到下一个终止密码子处才能终止翻译。这使翻译的产物比正常的蛋白质 C 端长出一段，可能会使蛋白质的功能改变或丧失，对细胞产生损害，严重的可致死。琥珀型抑制型 tRNA 的这种效应较强，约为 10%～15%，因此大肠杆菌中该终止密码子的使用频率最低。赭石型抑制很难分离到，在 10%以下，突变生长很慢，说明对 UAA 和 UAG 的抑制对大肠杆菌损伤较大。UAA 的使用频率较高。UGA 的使用频率最低，正常时被 Trp-tRNA 错读的概率为 1%～3%。

一个基因的错义突变抑制型 tRNA 会引起另一个基因的突变，从而引起蛋白质功能的改变。这对细胞来讲是一个进退两难的事情，既要抑制突变，又要防止新突变的发生。所以，没有发现很强的错义突变抑制型 tRNA，因为它能引起强烈的突变。抑制型 tRNA 突变的产生在使其识别突变的密码子的同时，有时会丧失识别原来密码子的功能。因此，琥珀型抑制突变体较多，因为细胞内识别一种密码子的 tRNA 不只一种。一种 tRNA 突变后，还有其他未突变的 tRNA 在起作用。如果只有一种 tRNA，该 tRNA 的突变将是致死的。

三、tRNA 可影响可读框

mRNA 的可译框架通常不发生变化，由起始密码子开始，至终止密码子结束。mRNA 分子中一个碱基的插入或缺失突变会使可读框变化，产生完全不同的翻译产物。移码抑制(frameshift suppression)可恢复正确的可读框架。这种抑制作用可发生在基因水平上，也可由基因外的因素(tRNA)参与完成。

在由同一碱基组成的密码子插入同一个碱基，如 GGG 变为 GGGG，是 tRNA 的移码抑制作用的最简单的例子。这种突变可由 Gly-tRNA 来抑制，因为其反密码子由 CCC 变为 CCCC，正好可识别 GGGG。这样，密码子和反密码子由三联体变为四联体。有时，移码突变抑制 tRNA 可识别两个以上的 4 碱基密码子。如细菌的一种抑制型 tRNA(Lys-tRNA)可同时识别 AAAA 和 AAAU，而正常时识别 AAA。另一个抑制型 tRNA 可识别 ACCN，N 为任何一种碱基。这种情况下，第四个碱基并不参与和密码子的配对，可能只是受到空间阻碍，而不影响下一个密码子的识别。在酵母线粒体中也发现了 tRNA 的移码突变抑制作用，一串 5 个 T 中插入或缺失一个 T 的突变均受到抑制。这可能是因为识别 UUU 的 Phe-tRNA 允许核糖体在任何方向上漏掉一个碱基。

在有些噬菌体和病毒中，移码是一种正常的现象，可造成翻译的中止或延长。如在 MS2 噬菌体，移码可使核糖体识别终止密码子，引起翻译的终止，然后识别裂解基因的起始密码子并翻译该基因。在无移码的情况下，核糖体继续翻译，不能识别裂解基因的起始

密码子，该基因不能表达。在反转录病毒，一条 mRNA 上存在不同的可译框架，移码可造成对不同阅读框的翻译，使不同的基因表达。使酿酒酵母的 Ty 转座子中，*TYB* 基因的表达必须经移码越过 *TYA* 基因的终止密码子。这表明正常的翻译过程也需要错读，只是较少见。这类移码有以下特征：

1）存在一不稳定序列(slippery sequence)使氨基酰 tRNA 与密码子配对，然后向前或向后移动一个碱基，移动后的密码子仍可和反密码子配对。

2）核糖体在移码位点暂停，使得氨基酰 tRNA 有充足的时间完成碱基配对的调整。引起核糖体暂停的原因可以是相邻的密码子所需的氨基酰 tRNA 为稀有 tRNA，或是由释放因子缓慢识别终止密码子，或是 mRNA 结构(如假结)的阻止作用。

已发现一些不稳定的序列。如 CUUAGGC 通常由 Leu-tRNA 识别 CUU，Arg-tRNA 识别 AGC。但 Arg-tRNA 含量很少，会引起核糖体的暂停，这时 Leu-tRNA 由 CUU 滑动一个碱基至 UAA，从而引起了移码，使下一个密码子 GGC 由 Gly-tRNA 识别。这种现象发生在 P 位，这时 Leu tRNA 已变为肽酰 tRNA，携带了新生链。tRNA 的移码可以向前，也可以向后。向前移动一个碱基可引起正 1 突变，向后移动一个碱基可引负 1 移码。

在体外实验中，翻译体系中含有过量的某一种氨基酰 tRNA 可引起移码的发生。体内的某些移码(如反转录病毒)依赖于非正常密码子的使用或 mRNA 分子内的二级结构。

第三节　核糖体的结构[18~23]

核糖体(ribosome)是翻译进行的场所，提供蛋白质合成过程中所需的各种生物活性。核糖体是细胞的重要器官。在一个生长旺盛的细菌中，约有 20 000 个核糖体。其蛋白成分占细菌总蛋白的 10%左右；RNA 成分占细菌总 RNA 的 80%左右。在真核细胞，相对比例小一些，但绝对含量更大。核糖体的数目与细胞蛋白质合成的活性密切相关。细菌的核糖体与 mRNA、mRNA 与 DNA 相连；真核细胞的核糖体与内质网或细胞骨架相连。核糖体由大小两个亚基组成。RNA 和蛋白质几乎都只有一个拷贝。RNA 和蛋白质相互作用，有的形成活性部位，有的提供构象，形成一个复杂而有序的结合体。

一、核糖体的构成

不同的细菌、真核生物胞质和细胞器的核糖体组成情况有很大差别。其共同特征是 RNA 含量比蛋白质高。核糖体中的蛋白质也可表示为 r-proteins。一个菌体细胞中的核糖体是相同的。大肠杆菌的核糖体研究的最为清楚，小亚基(30 S)由 16 S rRNA 和 21 种蛋白质组成。大亚基(50 S)由 23 S、5 S rRNA 和 31 种蛋白质组成。只有一种蛋白质有 4 个拷贝，其余的均为 1 个。真核生物细胞质核糖体比细菌的要大，RNA 和蛋白质的总含量多，RNA 分子长，蛋白质种类多。细胞器的核糖体与胞质有很大区别。最大的近似于细菌，含 70%的 RNA；最小的只有 60 S，RNA 含量少于 30%。细菌、叶绿体和真核生物细胞质核糖体的形状类似。细菌核糖体小亚基近似平台状，大亚基近似圆形。整个核糖体呈对称结构，大小亚基间有一定的缝隙，或称为通道(tunnel)。真核生物细胞质核糖体的结构与细菌相似。

mRNA 在核糖体中的定位还不是十分确定，较一致的看法是位于两个亚基的结合部。也有些模型认为 mRNA 像线一样穿过核糖的缝隙，但似乎应该位于较肤浅的部位。因为两个 tRNA 比核糖体还要大，插入核糖体内部的可能性不大，而位于其表而形成的空隙中的可能性较大。

核糖体中 RNA 的含量很高，几乎所有的蛋白质都与之结合存在。rRNA 主要形成核糖体的骨架，像一根线将蛋白质串起来，并决定蛋白质的定位。每种 rRNA 都有二级结构。在 rRNA 这样的大分子中，互补区域较复杂，单纯通过分析碱基配对区域并不能正确的预测二级结构形成的情况。分析 rRNA 形成二级结构情况的一种途径是比较相似物种间相应 rRNA 的序列，形成二级结构的部分比较保守，以形成相同的功能区。用该方法已确定了 16 S 和 23 S rRNA 的结构模型。16 S rRNA 的结构如图 6－3 所示。

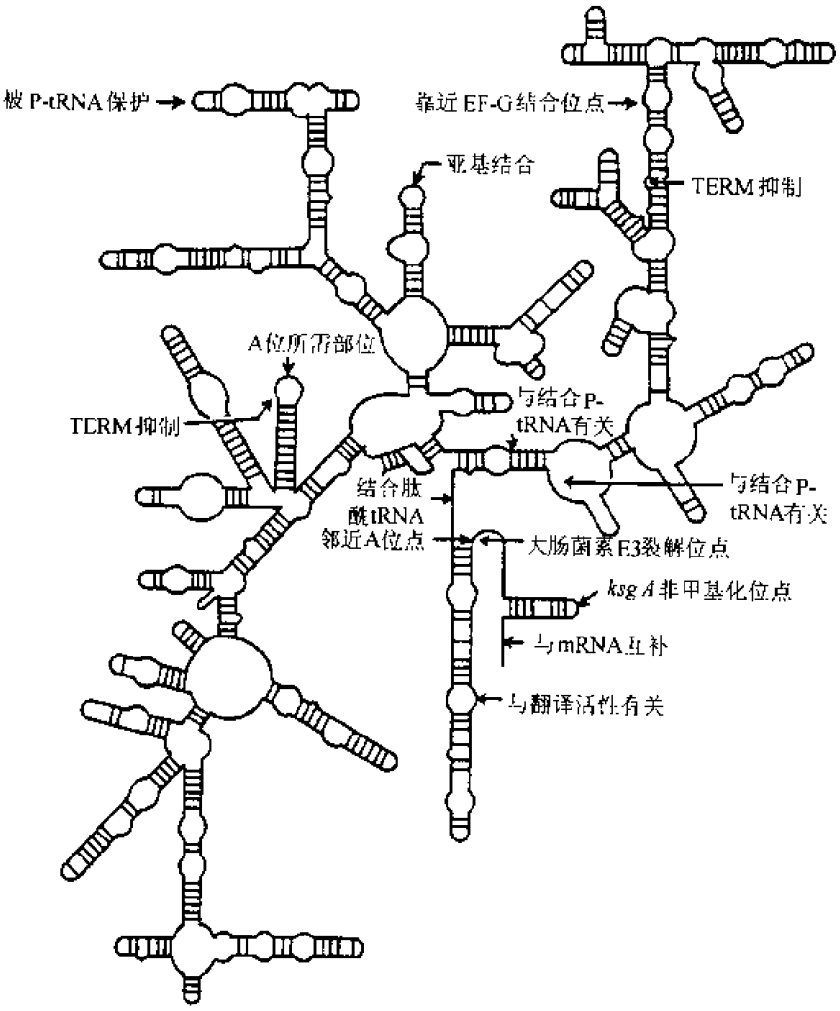


图 6－3　16 S RNA 的结构

在 16 S rRNA 中，形成了 4 个区域，有不足一半碱基呈配对状态。双螺旋区较短，小于 8 bp，并常含非配对碱基形成的泡状。线粒体和真核生物细胞质 rRNA 具有类似的结构。线粒体 rRNA 短些，功能区较少。真核生物细胞质 rRNA 较长，比细菌 rRNA 长出的

部分形成了新的功能区，因此功能区较多。经过 3-乙氧-2-氧丁醛(kethoxal)(可修饰未配对的 G)反应、补骨脂素(psoralen)交联反应、核酸酶优先降解位点的确定和 rRNA 与蛋白质的连接反应等实验，结果和上述模型相符，但形成的碱基对较少。

rRNA 在核糖体中的结构并不是一成不变的，mRNA 和 tRNA 进入核糖体、亚基的解离能引起 rRNA 构象的变化。16 S rRNA 在 30 S 亚基和 50 S 亚基中的反应性有区别，在游离状态和蛋白质合成过程中的核糖体中也有区别。还不清楚这种变化是由 mRNA 和 tRNA 与 rRNA 作用直接引起，还是由核糖体结构的变化引起。这说明核糖体的结构在蛋白质合成过程中有一定的灵活性。

甲基化修饰是 rRNA 的一大特征，特别在保守区域内。16 S rRNA 有 10 个甲基化修饰，集中在 3′端，23 S rRNA 有 20 个。真核生物的 18 S 和 28 S rRNA 各有 43 和 74 个，比原核生物多 3 倍。

除线粒体外，其他的核糖体都有 5 S rRNA。无论是原核生物还是真核生物，5 S rRNA 都各自有保守序列，特别是在与蛋白质结合的区域。5 S rRNA 富含相互配对的碱基，以此预测二级结构更加困难，现已提出了 20 种模型。在真核生物中，大亚基还有 5.8 S rRNA，序列近似于原核生物的 23 SrRNA 的 5′端。

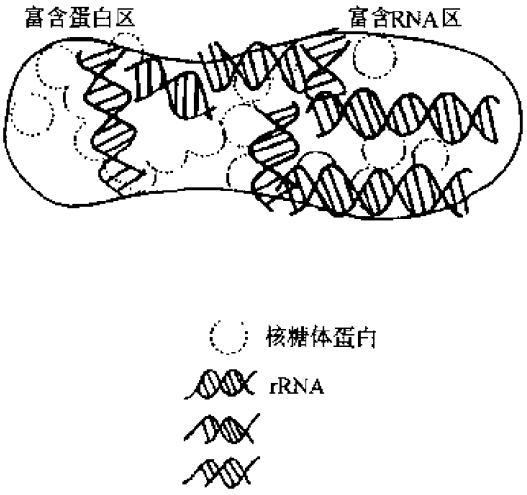


图 6 4 30 S亚基中各种成分的定位

rRNA 分子中有与蛋白质特异结合的位点。用保护实验可确定这些位点，它们都位于形成发夹结构的位置。有些核糖体蛋白与游离的 rRNA 就有很强的结合力。有些不能与游离的 rRNA 结合，只有 rRNA 与其他的蛋白质结合后，才有与之结合的能力。可能是蛋白质的结合会引起 rRNA 构象的变化。大肠杆菌中，所有的核糖体蛋白质都有与 rRNA 结合的能力。免疫电镜、交联反应、蛋白质与 rRNA 的相互作用和中子散射等技术可确定各种蛋白质在核糖体中的定位。现已明确了 30 S 亚基中蛋白质和 rRNA 的定位情况。

二、核糖体的装配

核糖体的装配过程中，涉及到蛋白质的某些基团与 rRNA 结合。大肠杆菌核糖体的各种组份可在体外装配成完整的核糖体，而且可进行装配反应的逆反应。经 CsCl 离心，细菌的核糖体有一部分脱落蛋白质(split protein)解离下来，30 S 和 50 S 亚基分别变成 23 S 和 42 S，称为颗粒核心(particle core)。在 LiCl 或 CsCl 溶液中，也有部分蛋白质按一定次序解离下来，说明这些蛋白质间有相互作用。除去 CsCl 或加入镁离子，这些蛋白质又可重新装配，形成有生物学活性的核糖体亚基。

30 S 亚基装配的第一阶段是 16 S rRNA 和 15 种蛋白质在低温下形成 RI 颗粒(RI

particle)，然后加热到 40℃，经分子重排为 21 S 颗粒，最后再结合 6 种蛋白质。EDTA 处理 30 S 亚基也能产生类似的 RI 颗粒。络合剂可去除镁离子，破坏亚基的空间结构，解离一些蛋白质。将残余颗粒加热后才能与失去的蛋白质重新结合。加热的作用是使 RI 颗粒形成更加紧密的结构，而这种结构又是体外装配所必需的。

在一些亚基装配缺失突变体(subunit assembly defective, *sad*)，会有装配前体堆积。前体颗粒含有 rRNA，但只结合了部分蛋白质。进一步分析，共有三种前体物质，21 S 为 30 S 亚基的前体，32 S 和 43 S 为 50 S 亚基的前体。体内的 21 S 前体与 RI 颗粒有相似的蛋白质成分，这些蛋白质在 CsCl 或 LiCl 处理时属最后解离的部分。可能它们最先装配入亚基的，所以最难于被去除。体内的前体含有前体 rRNA，RI 颗粒含成熟的 16 S rRNA。前体 rRNA 比成熟的稍长些，甲基化程度低些，构象会有些不同，影响与蛋白质的结合。

在缺乏某种或几种蛋白质的情况下进行装配实验，可确定每一种成分在装配中的作用，及对其他蛋白质装配的影响。经体外实验确定了 30 S 亚基的装配图，各种蛋白质以不同的次序与 rRNA 结合。

在 rRNA 转录时，装配即已开始。多数蛋白质的强结合位点位于 rRNA 先转录出的 5′端。在转录完成时，已结合了几种核糖体蛋白质。

三、核糖体的活性位点

核糖体可分为翻译区域(translational domain)和逐出区域(exit domain)，有 A、P、肽酰转移酶活性位点、EF-Tu、EF-G、5 SrRNA、mRNA 和逐出(E 位)等活性位点，如图 6-5 所示。翻译区域占 2/3，逐出区域占 1/3，与膜系统结合。

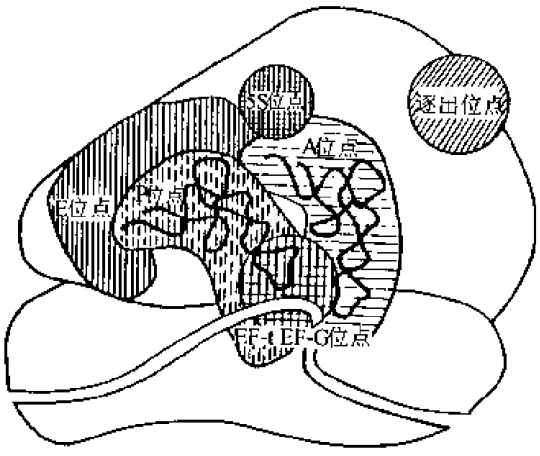


图 6-5　核糖体的活性位点

用标记的 tRNA 类似物与核糖体结合，然后观察核糖体的哪一种成分与之结合，可用于确定核糖体的活性位点，这种方法称为亲合标记实验。另一种方法是对亚基进行修饰，通过修饰使部分蛋白质或 rRNA 失去活性，看其对核糖体的功能有何影响。如 3-乙氧-2-氧丁醛能修饰 rRNA 的鸟嘌呤；四硝基甲烷可使蛋白质硝基化。但要将修饰的程度控

制在一定的范围内。

30 S亚基与mRNA的结合需要S1蛋白。S1蛋白对单链核酸有很强的亲合力，并能阻止mRNA形成分子内的二级结构，以利于翻译的进行。S1蛋白呈长链状，因而分析其在核糖体的定位较困难。实验证实，它和S18与S21组成一个活性区域，位于小亚基的峡部，参与起始tRNA识别AUG。与mRNA起始点碱基配对的rRNA的3′端也位于该区域。起始因子也结合在该区域中。如IF3可与16 S rRNA的3′端和一些参与mRNA结合的蛋白质发生交联。如果与大亚基结合的区域也位于该区域，IF3的结合可稳定mRNA-30 S复合物，而后被大亚基所取代。大小亚基的结合机制还不太清楚。位于亚基表面的16 S rRNA第791位碱基的突变影响大小亚基的结合。16 S rRNA的3′端区域与23 S rRNA有一互补区，对大小亚基的结合起一定作用。

16 S rRNA部分位于A位，它在1392至1407有一保守区，称为1400区(1400 region)，该区域参与和tRNA的结合。交联和保护实验证明肽酰-tRNA和氨基酰-tRNA均与这一区域结合。目前已鉴定出分别有几种蛋白质构成A和P位，但还不清楚它们间的作用机制。转肽酶活性位点已被嘌呤霉素定位，该位点包括了23 S rRNA和几种蛋白质。生长的肽链由50 S亚基伸出，距转肽酶活性位点约15 nm，呈伸展状态，当离开E位后发生折叠。L7与L12结构极其相似，只是L7的N端有一乙酰基。L7/L12形成了大亚基的柄部，功能是刺激有GTP酶活性的蛋白质发挥作用。EF G位点位于50 S亚基，距30 S亚基的S12(一种mRNA结合蛋白)很近，因此该位点分布在两个亚基上。保护实验证明，EF-G可保护23 S rRNA上两处碱基，其中一处(2660环)也被EF-Tu保护，说明EF-G位点和EF-Tu位点连接在一起。28 S rRNA也有类似的环，对该环的破坏引起肽链延伸的终止，说明这一环状结构的重要作用。

A位和P位要同时结合tRNA，每个tRNA结合一个密码子，所以二者相距很近。只有P位结合了肽酰-tRNA后，氨基酰-tRNA才能与A位结合，可见P位影响A位的活性。A位和P位的tRNA呈何种构象仍难于解释，tRNA分子的直径为2 nm，因此两个反密码子的距离不能超过1 nm。位于A位和P位的tRNA可能有相同的构象。mRNA在两个密码子间形成一定的角度，两个tRNA从不同的方向与mRNA结合。这种模式可解决A位和P位的立体结构与tRNA大小之间的矛盾。核糖体的移位可能通过角度的变化将tRNA由A位移动到P位，这样可使mRNA相对于核糖体的移动由tRNA来完成，以便于由密码子和反密码子间的配对来确定移动的距离。

核糖体上活性位点之间有着密切的联系，一个位点活性的变化可影响其他位点的功能。

四、核糖体在蛋白质合成中的作用

以前，人们认为核糖体是由多种具有催化功能的蛋白质通过蛋白质间及和rRNA间的相互作用装配起来的。自发现RNA有催化活性后，使人们想到RNA在核糖体中可能有更加重要的功能。现已发现在翻译的各阶段均有rRNA的参与，rRNA的正确结构是蛋白质发挥催化功能所必需的。在翻译起始过程中，rRNA的3′端直接与mRNA作用；16 S rRNA在A和P位直接与反密码子作用；23 S rRNA与肽酰-tRNA的CCA端直接

作用;大小亚基的结合可能涉及到16 S和23 S rRNA间的碱基配对关系。在翻译的各个阶段,抑制翻译的抗生素几乎都作用于rRNA,说明rRNA的重要性。

结构分析、碱基修饰和突变证明了16 S rRNA分子中有些碱基是核糖体的功能所必需的。如用大肠菌素(colicin) E3切除16 S rRNA的3′端50个核苷酸,可丧失核糖体结合IF3、识别mRNA和结合tRNA的能力,完全阻止翻译的起始。在16 S rRNA的3′端,有--6碱基的序列,与mRNA中AUG上游的6个碱基相互补,使二者互补性下降的突变能中止翻译,使互补性增加的突变可增强翻译的效率。这说明rRNA和mRNA间的互补对翻译的起始有重要作用。

rRNA的突变影响翻译的特异性。如16 S rRNA的3′主要功能区的突变可抑制翻译的终止;rRNA与mRNA间的识别作用能引起移码,越过mRNA上的一个碱基,这表明在核糖体的移动过程中,rRNA与mRNA可能发生直接的碱基识别。通过分析16 S rRNA受保护的碱基位点,发现正在进行翻译的核糖体中16 S rRNA的结构会发生一些变化,这时16 S rRNA受保护的碱基位点分布在整个分子中(一级结构),但在三级结构中,这些位点集中在一定的区域,执行特定的功能。引起16 S rRNA结构变化的因素有50 S亚基的结合、mRNA及tRNA与核糖体的结合。这说明在翻译过程中的不同阶段,核糖体的构象发生变化,使不同的rRNA发挥作用。上已提及,A位和P位均有16 S rRNA的参与。结合在A位的tRNA可保护16 S rRNA分子上的两个区域,其中一个位于3′端,另一个称为530环。530环处的突变可抑制UGA、UAG和UAA的识别。受P位rRNA保护的碱基位点在一级结构中很分散,但在三级结构中相对集中。16 S rRNA的1400区能与肽酰-tRNA发生交联,说明这一区域是P位的组成部分。tRNA的这些效应是由反密码子茎环结构引起的(见图6-6)。

对23 S rRNA了解没有16 S详细,但也鉴定了一些功能位点。它也参与了A位和P位的形成。tRNA3′CCA端的一段序列能保护23 S rRNA的几个碱基,说明肽酰-tRNA在P位通过3′CCA端与23 S rRNA作用。E位也可与tRNA结合,空载的tRNA经过该位点离开核糖体。23 S rRNA参与形成E位点。肽酰转移酶活性由哪种成分来提供的还不清楚,但rRNA肯定有重要的作用,有些rRNA的突变可抵抗某些抗生素对转肽酶活性的抑制作用。目前还未能在核糖体中发现有催化活性的蛋白质。将50 S亚基中大多数蛋白质成分抽提去除,所残余的23 S rRNA及与之相连的少量蛋白成分具有转肽酶活性。破坏rRNA的二级结构后,转肽酶活性消失。由于目前还不能制备有完整二级结构而不含蛋白质成分的23 S rRNA,所以还不能肯定的说23 S rRNA具有转肽酶的活性。可以推测,50 S亚基中的某些蛋白成分使23 S rRNA形成特定的空间结构,才能形成转肽酶活性。这和RNA剪接过程中RNA的剪切活性有很大相似。

核糖体在细胞内除参与翻译过程外,一部分以游离状态存在。在处于生长状态的细菌,游离的核糖体约占20%。

真核生物的核糖体的功能还知之甚少。其rRNA的作用可能类似于原核生物。一些抑制核糖体功能的物质都是通过与rRNA作用的。如蓖麻毒蛋白凝集素(lectin ricin)就是使28 S rRNA脱去一个腺嘌呤而抑制翻译的。

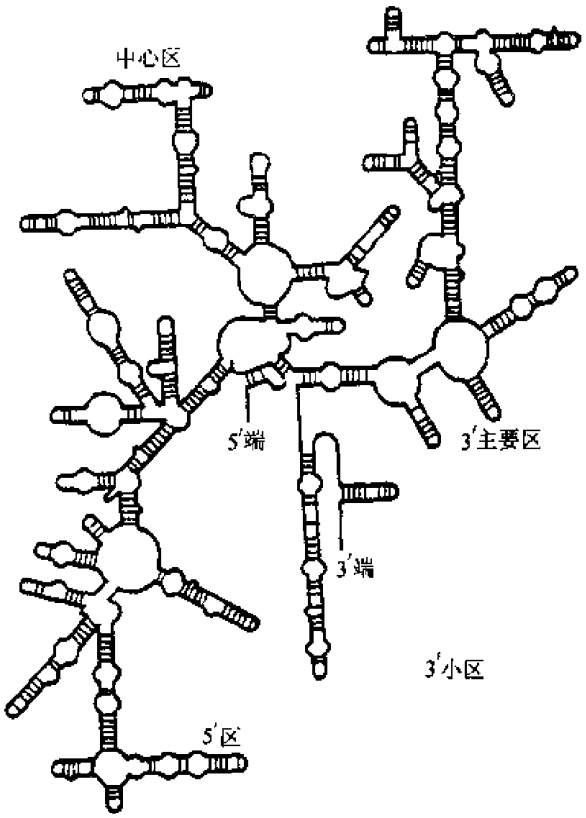


图 6-6　16 S rRNA 分子中的功能区域

第四节　原核生物的翻译过程[24~28]

无论原核生物还是真核生物，翻译可分为三个阶段，即起始(initiation)、延伸(elongation)和终止(termination)。翻译的速度很快，在 37℃ 为 15 个氨基酸/秒。一个 300 个氨基酸组成的蛋白质分子只需 20 秒就能完成。

研究翻译过程的方法主要是应用体外翻译系统。体外翻译系统的速度比体内慢一个数量级。下面讲述原核生物翻译的过程。

一、起　　始

翻译的起始是核糖体的大小亚基、tRNA 和 mRNA 在起始因子的协助下组合成 70 S 起始复合物的过程。参与起始过程的核糖体亚基由游离存在的和终止后核糖体解离产生的亚基提供。从这一点来看，起始和终止是密切相关的。

1. 起始的第一步是 30 S 亚基与 mRNA 的结合

在 mRNA 编码区的上游有一段短的序列能为 30 S 亚基识别，称为核糖体结合位点(ribosome-binding site)。30 S 亚基单独并不能和 mRNA 结合，必须有起始因子(initiation

factor,IF)参与。起始因子发现于 30 S 亚基"洗涤"实验。用氯化氨处理过的 30 S 亚基在体外不能与 mRMA 结合,后来发现有些与 30 S 亚基结合疏松的蛋白因子被洗下来,被鉴定为起始因子。起始因子只存在于游离的 30 S 亚基中,每种起始因子的拷贝数比游离 30 S 亚基少得多。细菌中有三种起始因子,分别为 IF1、IF2 和 IF3。其中 IF3 促进 30 S 亚基与 mRNA 的结合;IF2 参与起始 tRNA 与 30 S 亚基的结合;IF1 的功能还不特别确定,它与 30 S 亚基结合只是作为起始复合物的一个组分,可能只起稳定作用,而不是识别 30 S 亚基中的特别成分。

IF3 有两个作用,第一是稳定游离的 30 S 亚基。当 30 S 亚基从核糖体解离后,要依靠 IF3 来维持游离状态。这种作用对维持核糖体与游离亚基间的平衡有重要作用。IF3 结合 30 S 亚基可阻止与 50 S 亚基缔合成完整的核糖体,因此 IF3 又称为抗缔合因子。一个 IF3 可结合一个 30 S 亚基。细胞内 IF3 的数目比 30 S 亚基少一些。IF3 的第二个功能是保证小亚基与 mRNA 的结合,在无 IF3 时,小亚基不能与 mRNA 形成起始复合物。但小亚基有选择 mRNA 上转录起始点的功能,所以只有 IF3 和小亚基共同作用,才能形成正确的起始复合物。

mRNA 与 30 S 亚基的结合需要 mRNA 与 16 S rRNA 间的碱基配对关系。受核糖体保护的 mRNA 约有 30 个碱基长,但 AUG 或 GUG 不在保护区内。在 AUG 上游 10 个碱基处有一保守区域,其一致序列为 AGGAGG(S-D 序列)。在 16 S rRNA 的 3′端区域存在与之互补的保守序列 CCUCCU。该序列的点突变可抑制翻译的进行。在真核生物中,rRNA 的 3′端区域也存在一类似的区域,为 CCUCC。但没有在 mRNA 中发现其互补序列。

2. 30 S-IF3 - mRNA 复合物与起始 tRNA 结合

在 mRNA 的核糖体结合位点,有一段序列是可译框架起始点的标志,即起始密码子,通常为 AUG,细菌中有时也用 GUG 和 UUG。AUG 是甲硫氨酸的密码子,甲硫氨酸有两种 tRNA,一种识别起始 AUG,一种识别延伸过程中的 AUG。在细菌和真核生物的细胞器,起始 tRNA 携带的是甲酰化的甲硫氨酸(fMet-tRNAf)。起始 tRNA 先由甲硫氨酸与 tRNAf 结合形成 Met-tRNAf,再由 10-甲酰四氢叶酸为供体,将甲硫氨酸的 NH_2 甲酰化,形成 fMet-tRNAf。这种"封闭"使起始 tRNA 携带的氨基酸丧失了形成肽键的能力。内部的 AUG 由 Met-tRNA 识别。起始 tRNA 可识别 AUG、GUG 和 UUG,但三种起始密码子的起始效率不同。用 GUG 替代 AUG,起始效率下降一半;用 UUG 替代 GUG,起始效率又下降一半。

AUG 和 GUG 作为起始密码子,代表甲酰甲硫氨酸;位于内部时,AUG 代表甲硫氨酸,GUG 代表缬氨酸。这种意义的差异是由核糖体决定的。在 30 S - IF3 - mRNA 复合物中,起始密码子正好位于 P 位,而且只有起始 tRNA 才能进入该复合物的 P 位,其他的 tRNA 只能进入 70 S 起始复合物的 A 位。因此,只有位于核糖体结合位点中的 AUG 或 GUG 才能成为起始密码子。

作为起始 tRNA 的 $fMet\text{-}tRNA_f$,在结构上和 $Met\text{-}tRNA_m$ 及其他 tRNA 是不同的。起始 tRNA 所负荷的为甲酰甲硫氨酸,没有游离的氨基,不能形成肽键。但这一特征并不是绝对的,因为在真核生物,起始 tRNA 为 Met-tRNA。起始 tRNA 还有另外两个重要的特

征(如图 6－7 所示)。一是接受茎最后两个碱基在所有的 tRNA 中均呈配对状态,起始 tRNA 是唯一的例外。如将这两个碱基突变成配对状态,起始 tRNA 就可参与延伸过程,说明这种非配对状态阻止了起始 tRNA 被用于延伸。另一个特征是起始 tRNA 反密码茎中有三个 GC 对。这几个碱基对的突变能使起始 tRNA 不能进入 P 位。

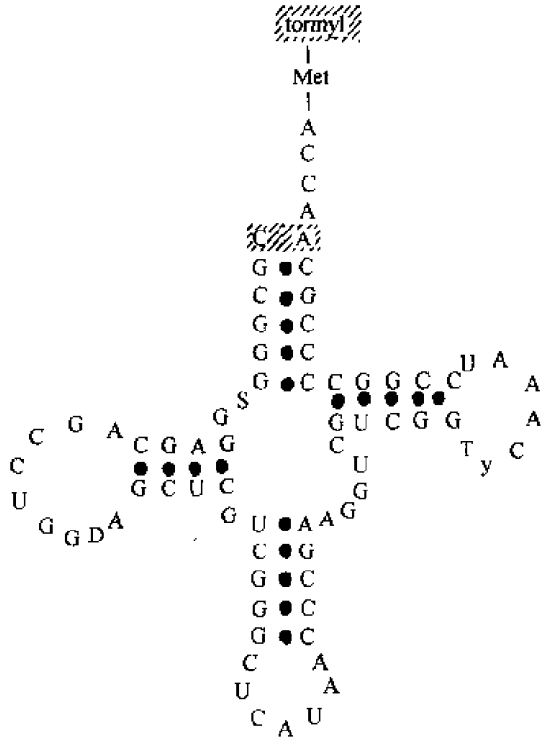


图 6－7　起始 tRNA 的结构特征

起始 tRNA 进入 P 位是在 IF2 的协助下完成的。IF2 先与起始 tRNA 形成二元复合物(IF2－fMet-tRNAf),IF2 只能与起始 tRNA 相互作用,因此保证了其他的 tRNA 不能用于起始。另外,IF2 还有依赖核糖体的 GTP 酶的活性。二元复合物具备了与 30 S 亚基结合的能力。同时,GTP 也加入到复合物中去。

3. 70 S 起始复合物的形成

30 S-IF3－mRNA-IF2－fMet-tRNAf-GTP 复合物有与 50 S 亚基结合的能力。50 S 亚基的结合,可使 IF3 游离,看来 IF3 与 50 S 亚基间的竞争决定了核糖体的工作状态。50 S 亚基的结合还激活了 IF2 的 GTP 酶活性,水解 GTP 释放能量,并解离下来。这样就形成了 70 S 起始复合物(30 S-mRNA-fMet-tRNA－50 S)。这时所有的起始因子都解离下来而不参与延伸过程,因此它们只起辅助因子的作用。

起始甲酰甲硫氨酸的甲酰基在肽链合成至 15～30 个氨基酸时被去除。在细菌和线粒体,有一种特殊的去甲酰基酶。如果所合成的蛋白质第一个氨基酸不是甲硫氨酸,在氨肽酶的作用下,甲硫氨酸也被去除。这两个反应几乎是同时进行的。

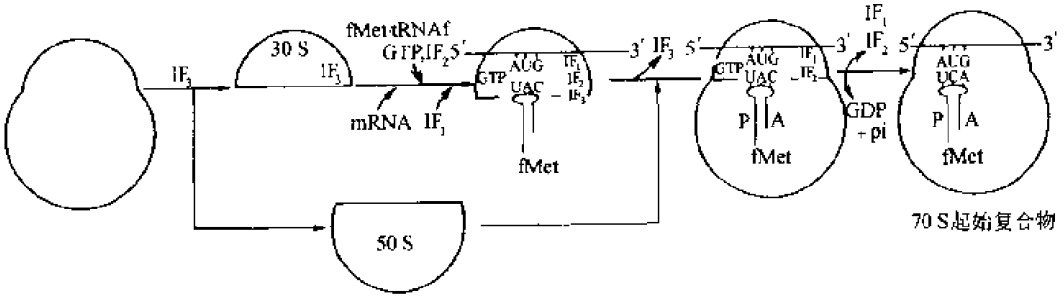


图 6-8　70 S复合物形成过程

二、肽链的延伸

肽链的延伸以氨基酰- tRNA 进入 70 S 起始复合物(进位)的 A 位为标志。除起始 tRNA 外,其他的氨基酰- tRNA 均有进入 A 位的能力。这一过程需要延伸因子(elongation factor,EF)参与,在细菌为 EF-Tu。EF-Tu 只有在氨基酰 tRNA 进位时才能与核糖体结合,进位完成后便从核糖体上解离下来,参与下一个进位过程,是一个典型的辅助因子。

EF-Tu 与鸟苷酸结合,其活性受鸟苷酸的调节。有多种蛋白质通过这种方式来调节自身的活性。在 GTP 存在时,EF-Tu 有活性;当 GTP 水解为 GDP 时,呈无活性状态;当 GDP 被 GTP 取代后,又恢复活性。EF-Tu-GTP 能结合氨基酰- tRNA 形成 EF-Tu-GTP-氨基酰- tRNA 三元复合物,而不能与起始 tRNA 结合,这保证了起始 tRNA 不参与肽链的延伸。在核糖体的 P 位被起始 tRNA 或肽酰- tRNA 占据的情况下,三元复合物进入到 A 位,GTP 水解,并释放 EF-Tu-GDP。而后发生转肽反应,即 A 位的氨基酰- tRNA 所携带的氨基酸与 P 位的肽酰 tRNA 的肽链(或起始氨基酸)形成肽键,肽链延长了一个氨基酸。催化该反应的酶为肽酰转移酶(peptidyl transferase)。EF-Tu-GDP 在 EF-Ts 的作用下形成 EF-Tu-EF-Ts(也称为 T 因子),释放 GDP。EF-Ts 被 GTP 取代,形成 EF-Tu GTP。这一过程称为 Tu-Ts 循环。

一个菌体细胞内有大约 70 000 个 EF-Tu 分子,与氨基酰- tRNA 的数目相似。这表明可能所有的氨基酰- tRNA 都以三元复合物的形式存在。EF-Tu 由两个基因编码,占菌体蛋白的 5%,多数以二元或三元复合物存在。一个菌体中有 10 000 个 EF-Ts,与核糖体的数目相似。

用 GTP 的类似物 GMP-PCP(不能水解供能)代替 GTP 与 EF-Tu 结合,可使氨基酰-tRNA进入 A 位,但不能形成肽键。黄色霉素(kirromycin)可抑制 EF-Tu 的活性。当黄色霉素与 EF-Tu 结合时,EF-Tu-GTP 仍能使氨基酰- tRNA 进入 A 位,但不能释放 EF Tu-GDP,使肽键不能形成,蛋白质合成停止。这说明蛋白质合成过程中,抑制一个步骤可阻止下一个步骤的进行。EF-Tu-GDP 的不能释放造成了核糖体不能形成正确的构象,因而不能完成转肽反应。

转肽反应使 P 位的肽酰- tRNA 的肽链转移给 A 位的 tRNA,而 P 位的 tRNA 变为空

载。转肽酶活性位于核糖体的大亚基上，该位点可将 A 位和 P 位的 tRNA 拉近，以便形成肽键(图 6－10)。完成这一反应需要 rRNA 和 50 S亚基的蛋白质，但起催化作用的更像是 rRNA。

转肽结束后，核糖体沿 mRNA 向前移动一个密码子的距离，这一过程称为移位(translocation)。移位的同时将 P 位的空载 tRNA 经 E 位逐出核糖体，A 位的 tRNA 移至 P 位，以便新的氨基酰进入。移位可分为两个阶段。第一阶段，当肽键形成后，A 位的 tRNA 或是经自身移动，或是 50 S 亚基相对于 30 S 亚基移动，进入 P 位；第二阶段，tRNA 的反密码子端移动至 P 位，并与密码子配对(图 6－11)。

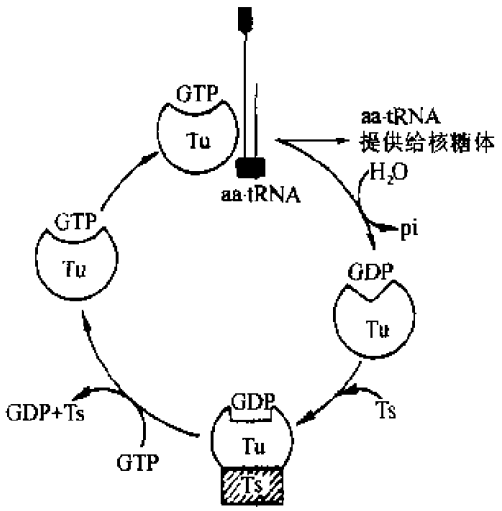


图 6－9　EF-Tu-Ts 循环

移位需要 GTP 和另一种延伸因子 EF-G。

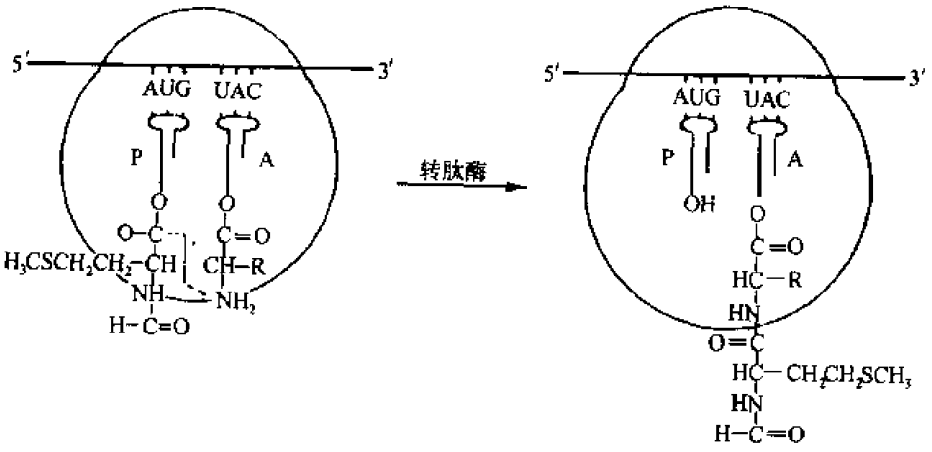


图 6－10　肽键的形成

EF-G 在菌体内的数目与核糖体基本相同(每个细胞约 2000 个)。EF-G 与核糖体结合可刺激移位，GTP 水解后解离下来。在 GTP 的类似物(GMP-PCP)存在时，EF-G 也能与核糖体结合，但不能发生移位。GTP 的水解是核糖体的功能，EF-G 不能催化 GTP 的水解。在 EF-G 不存在的情况下，核糖体也可以发生移位反应，但效率很低。这说明移位是核糖体固有的性质，但可被 EF-G 所催化。

参与蛋白质合成还有一种 4.5 S RNA，长 114 个碱基，在大肠杆菌中多呈双链状态，一个菌体细胞中有约 3000 个分子，只有少部分与核糖体结合。4.5 S RNA 基因的突变可抑制细菌的生长，该突变可被 EF-G 基因的突变所抑制，说明它和 EF-G 相互作用。部分分泌型蛋白质的合成需要 4.5 S RNA。

EF-Tu 和 EF-G 不能同时与核糖体结合，二者交替与核糖体结合和释放。在梭链孢酸(fusidic acid)存在时，EF-G 能与核糖体结合，GTP 也可水解，核糖体也可正常移位，但

EF-G-GDP 不能释放，使肽链的延伸终止。目前还不清楚两种延伸因子间的排斥是通过核糖体的构象引起还是二者竞争同一结合位点引起。二者交替与核糖体结合说明蛋白质合成是一个有序的过程。两种延伸因子与核糖体的结合都要有 GTP 而不是 GDP 的存在，可能是 GTP 能使延伸因子形成有活性的构象形式。

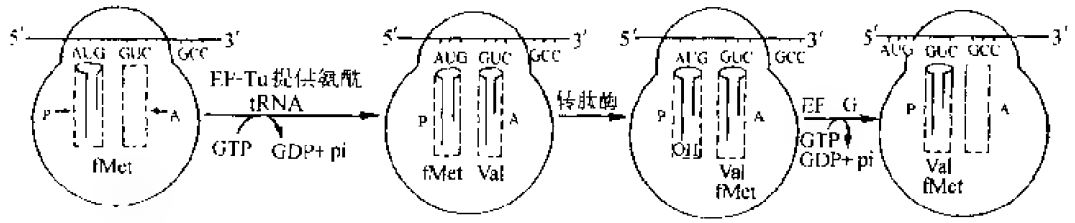


图 6－11　延伸的过程

三、合成的终止

有三种终止密码子，UAG 称为琥珀密码子，UAA 称为赭石密码子，UGA 称为乳石密码子。在细菌中，UAA 使用频率最高，UGA 次之，UAG 最低。没有一种 tRNA 能识别终止密码子，而是由蛋白质因子来识别。在大肠杆菌，有二种蛋白质因子催化终止反应，称为释放因子(relcasing factor，RF)。RF1 识别 UAA 和 UAG，RF2 识别 UGA 和 UAA，RF3 可刺激 RF1 和 RF2 的活性。RF 必须在 P 位有肽酰－tRNA 时，才能识别 A 位的终止密码子。释放因子的数量比起始因子和延伸因子少得多，每个细胞只有 600 个左右，相当于每 50 个核糖体一个释放因子。释放因子基因的突变可降低翻译终止的效率，表现为通读增强；过量的表达可增强终止的效率。这说明释放因子与误读终止密码子的 tRNA 间有竞争作用。

合成的终止包括肽链的释放、tRNA 的逐出和核糖体与 mRNA 的解离。新生肽链从 tRNA的解离类似于延伸过程中的转肽反应，而释放因子可能改变了转肽反应的性质，使肽链释放。核糖体与 mRNA 是如何解离的，还不清楚，可能由释放因子引起核糖体构象的变化，也可能有其他的蛋白质因子参与。

第五节　真核生物的蛋白质生物合成过程[29～34]

和原核生物一样，真核生物的蛋白质生物合成的过程包括起始、延伸和终止。但和原核生物比较起来，有很多特点。

一、合成起始

真核生物蛋白质生物合成起始过程中，需要甲硫氨酰－tRNAi、GTP、ATP 和十几种起始因子(eIF)参与。其具体过程如下。

1．80 S 核糖体的解离

80 S 亚基由 60 S 大亚基和 40 S 小亚基结合而成。正常生理情况下，80 S 核糖体和 60 S、40 S 大小亚基之间，存在一动态平衡，即：

$$80\ S \rightleftharpoons 60\ S + 40\ S$$

因此，即使在蛋白质合成十分旺盛时，也会存在少量的游离的 60 S 和 40 S 亚基。有三种核糖体亚基结合的起始因子（eIF－1A、eIF－3 和 eIF－3A）。多亚基因子 eIF－3 可与 40 S 亚基结合，形成 43 S 亚基，又称 43 Sn。eIF 1A 对 43 S 亚基有稳定作用。eIF－3 与 40 S 亚基的结合，干扰了 40 S 亚基与 60 S 亚基的结合。eIF－3 还可与 mRNA、rRNA 及 mRNA 的帽子结构相互作用。此外，eIF－3 可与 60 S 亚基结合从而阻止 60 S 和 40 S 亚基的结合。在三种起始因子的共同作用下，形成 43 S 亚基，是合成起始的开端。

2．三元复合物和 43 S 前起始复合物的形成

Met-tRNAi 可识别起始密码子 AUG。Met-tRNAi 首先与 GTP 和 eIF－2 形成三元复合物，而后与 43 S 亚基形成前起始复合物。在起始过程结束后，eIF－2－GDP 复合物释放。GDP 须变换成 GTP，方能重新进入下一个起始过程。催化 GDP 和 GTP 间交换反应的是 eIF－2B，它是一种 GTP 结合蛋白，需辅酶 NADPH，NAD 和 NADP 可抑制其活性。eIF－2 分子中，β 亚基与 GTP 结合，同样，γ 亚基也具有与 GTP 结合的位点。eIF－2 结合 GDP 的能力比 GTP 高上百倍。eIF－2·GTP 复合物不稳定，与 Met-tRNAi 结合后，形成稳定的三元复合物，再与 43 Sn 结合形成 43 S 前起始复合物。该结合过程需 eIF－1A 和 eIF－3 参与。43 S 前起始复合物的形成不需要 GTP。eIF－2α 的磷酸化可影响 GDP/GTP 的交换反应，从而调控 eIF－2 的活性。

eIF－2 还有其他的一些功能。如可与 mRNA 结合；与起始点的选择有关；eIF－2α 和－2β 亚基含有与 RNA 结合的保守区域等。

3．43 S 前起始复合物与 mRNA 的结合

在结合 mRNA 前，有三种 eIF 共同作用，消除 mRNA 的 5′非翻译区域（5′untranslated region，5′UTR）的分子内二级结构，即 eIF－4F、eIF－4B 和 eIF－4A。

eIF－4F 由三个亚基组成，可识别 mRNA 的 5′帽子结构 $m_7G5'ppp5'X$，这一过程需 ATP 参与。在 eIF－4F 分子中，eIF 4α 亚基又称 eIF－4E，独自具有结合帽子结构的功能，另两个亚基 eIF 4β 和 eIF－4γ 不参与此过程。eIF－4E 在细胞核中与 mRNA 结合，当转移到细胞浆中后再与 eIF－4F 的其他亚基结合。eIF－4F 的 γ 亚基是糖蛋白，具有很多蛋白酶的识别序列和 RNA 识别区域。此亚基的功能还不十分清楚。在脊髓灰质炎病毒感染时，该亚基受到水解，具帽子结构的 mRNA 的翻译受到抑制。但应当指出，该亚基的降解并不一定导致翻译的抑制。因而，在脊髓灰质炎病毒的感染时，定存在其他的机制。eIF－4F 的 α 亚基与 eIF－4A 十分相似，甚至可以相互替代。它以游离的形式大量的存在于细胞中。小鼠中有两种 eIF 4A 的基因，分别编码 eIF－4AI 和 eIF－4AII。游离的 eIF－4A 与 eIF 4AI 相同，而 eIF－4F 的 α 亚基是 eIF－4AIeIF－4AII 以 4∶1 的比

列组合而成的。eIF－4A 分型的意义尚不清楚。eIF－4A 在 eIF4－B 的存在下可与帽子结构相互作用，属于“致死”框蛋白家族，有螺旋酶的活性；含有 ATP 结合位点和 ATPase 位点；有依赖 RNA 的 ATPase 活性和依赖 RNA 的螺旋酶活性，eIF－4F 也有后一种活性，且受 eIF－4B 的激活。

eIF－4B 含结合 RNA 的区域，在 eIF－4A 或－4F 及 ATP 的存在下，能与帽子结构和 RNA 相互作用。此外，尚有依赖核糖体的 ATP 酶的活性，eIF－4A 和 eIF－4F 能增强这种活性。

eIF－4F 的作用机制如下：在 eIF－4α 亚基及 ATP 存在下，eIF－4F 与 mRNA 的帽子结构结合，eIF－4A 结合上来，形成螺旋酶复合物，开始使 mRNA 的 5′UTR 的二级结构解旋。这一过程需 ATP 的水解。eIF－4B 可促进 eIF－4α 亚基的再循环。在细胞内 eIF－4A 多于－4B 和－4F 基础上，eIF－4A 可维持与 mRNA 帽子结构的结合状态，从而维持 5′UTR 的单链状态。在此基础上，43 S 前起始复合物与 mRNA 结合形成 48 S 前起始复合物。具体何种因子参与这一过程还不十分清楚。eIF－3 对核糖体和 mRNA 均有亲合力，且 eIF－4B 和－4F 对 eIF－3 具有亲和力，可以推测上述因子参与了 43 S 前复合物与 mRNA 的结合过程。

4．起始密码子 AUG 的选择

40 S 亚基与 mRNA 的结合并不能使 Met-tRNAi 正好与 AUG 结合。AUG 一般位于帽子结构下游的 50～100 个核苷酸。Kozak 提出了一种识别 AUG 的机制，已在很多的研究中得到了证实。43 S 前起始复合物包括 eIF－3 和 eIF－2·GTP·Met tRNA 由帽子结构处沿 5′UTR 移动，直至第一个 AUG 处。有关解旋因子 eIF－4A、－4B 和－4F 在此过程中的变化尚不清楚。鉴于 eIF-4B 与核糖体的作用及其依赖核糖体的 ATPase 的活性，该因子可能在此过程中起重要作用。

对 AUG 的识别可能是反密码子与密码子间的碱基配对。但还有其他的因素参与，因为对不同 AUG 的识别效率不同。无论真核生物还是原核生物，在 AUG 前后（～3 到＋4）有一保守序列 CCC(A/G)CCAUGG，同核糖体小亚基的 RNA(ssrRNA)的一保守序列互补性很强。特别是 ss rRNA 与 mRNA 中 AUG 上游的三个碱基间的稳定结合，造成了 43 S 复合物移动的缓慢，为 Met-tRNAi 识别 AUG 提供了时间。

除 Met-tRNAi 和 rRNA 外，还有其他一些因素在 AUG 的选择上都有重要的作用。eIF－2α 或 β－亚基的突变可致使选择 UUG 为起始密码子。eIF－2/mRNA 的比例较低可引起利用 3′远端的 AUG 为起始密码子。这说明 eIF－2 在起始密码子的正确选择上有重要作用。

5．48 S 起始复合物与 60 S 亚基结合

43 S 起始复合物在 AUG 处正确定位后，在 eIF－5 的作用下，eIF－2·GTP 水解为 eIF－2·GDP，从而导致 eIF－2·GDP 和其他的 eIFs 由 40 S 亚基上解离下来，60 S 亚基与 40 S 亚基结合形成 80 S 起始复合物。与 60 S 亚基结合的 eIF－3A 也释放。有些文献指出，eIF－2·GDP 仍结合在 80 S 复合物形成中，在 eIF2－B 的作用下才释放出来，并转换成 eIF－2·GTP。由于资料有限，尚无法定论。

80 S起始复合物形成后，第一个肽键的形成还需要 eIF－5A，它含有一个修饰的氨基酸，并与其功能有关。具体的机制还不清楚。

在细小病毒中，有一种不同的起始机制。该病毒的 5′UTR 非常长，在 610 至 1000nt 之间。并具有复杂的二级结构。它含有多至 15 个的 AUG 密码子，且种内和种间的位置、数目的保守性很差。这表明这些 AUG 不能被核糖体移动选择机制用为有效的起始密码子。它采用内部翻译起始的机制。在 5′UTR 中，有一内部起始所需的最小片段，称为内部核糖体进入片段(internal ribosome entry segment，IRES)。核糖体内部进入位点位于 IRES 的 3′端，长 25nt 左右，具有二级和三级结构，且对起始点的选择有重要作用。在脑心肌炎病毒的(EMCV)R 株，翻译起始于第十一个 AUG(834nt)，有时也起始于第 12 个 AUG(846nt)。第十位的 AUG(826nt)上下文序列具有类似于核糖体移动选择模型中起始部位的特点，但翻译却不由此起始。内部起始的过程中，也需要反式作用的蛋白因子参与。有关这一问题的详细论述，请参阅有关文献。

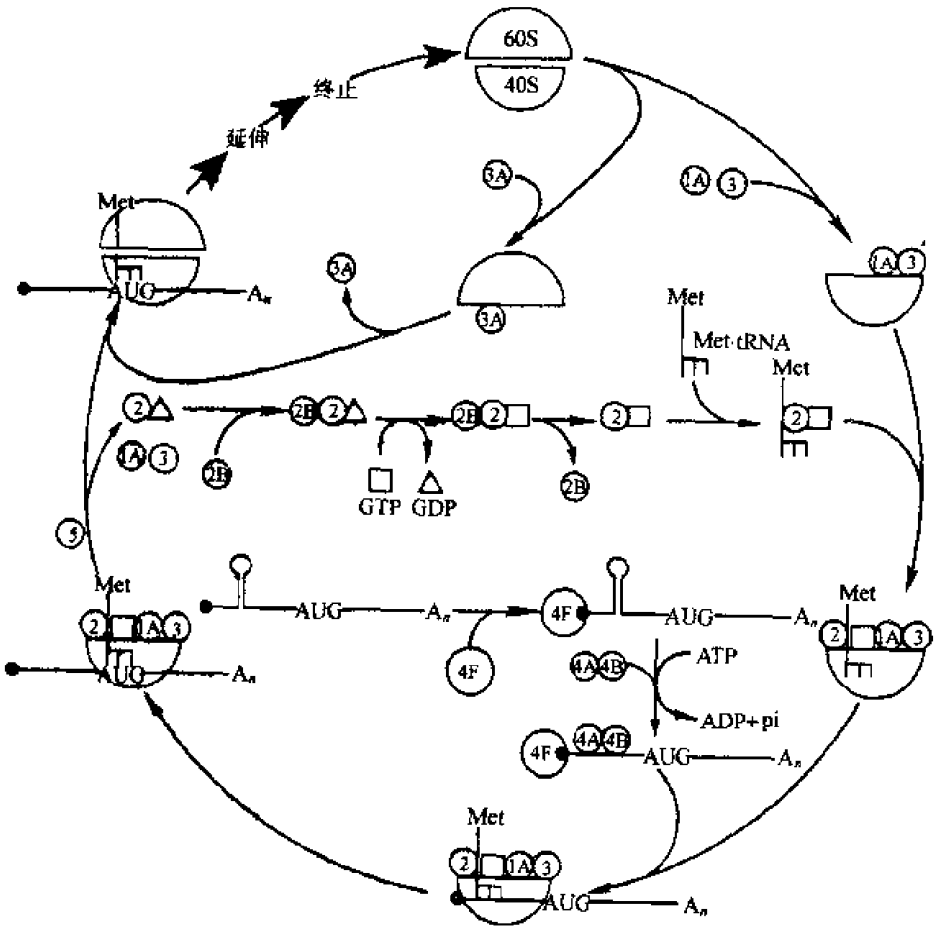


图 6－12　真核生物蛋白质生物合成的起始过程

二、肽链的延伸

肽链的延伸是将 mRNA 的核苷酸序列翻译为多肽链的氨酸序列的过程。其中，翻译的准确性是该过程的关键。

1. 延伸因子

在真核生物中，普遍存在着延伸因子 1(elongation factor - 1，eEF - 1)和 2(elongation factor - 2，eEF - 2)，在真菌中还有延伸因子 3(elongation factor - 3，eEF - 3)。

(1) 延伸因子 1　eEF - 1 是一种多聚蛋白质，主要负责将氨基酰- tRNA 转运至核糖体。大多数生物中由 4 种不同的亚基组成，分别为 α、β、δ 和 γ。α 亚基负责与氨基酰- tRNA 结合，并需 GTP 的参与。GTP 可能是 α 亚基的变构激活剂。GTP 同样是原核生物的 EF-Tu、IF - 2 及真核生物的 eIF - 2 所必需的，而这些因子的功能均是与氨基酰- tRNA 结合。α 亚基的 GTP 酶活性较低，但氨基酰- tRNA、核糖体和 mRNA 能增强这种活性。GTP 水解后 α 亚基从核糖体上解离下来，形成 eEF - 1α·GDP。eEF - 1α·GDP 必须转变为活性形式 eEF - 1α·GTP，才能继续发挥作用。β 亚基有核苷酸交换活性，能加速 eEF - 1α·GDP 转变为 eEF - 1α·GTP，类似于原核生物的 EF-Ts。γ 亚基具有增强 α 亚基的活性。δ 亚基的功能不十分确定，结构上与 β 亚基有同源性，特别是 C 端区域，而且 β 亚基的交换活性就在 C 端，因此推断 δ 亚基也可能具有核苷酸交换活性。但 N 端二者无同源性，而且不同来源的 δ 亚基的 N 端同源性也较差，因此 δ 亚基也可能有其他的功能。

eEF - 1 的 α、β 和 δ 亚基均可被蛋白激酶 C(PKC)磷酸化；α 亚基的修饰还包括甲基化和磷酸甘油乙醇胺基化。这些修饰的意义还不十分清楚，但肯定与 eEF - 1 的活性调控有关。

eEF - 1α 的过量表达在酵母可导致翻译忠实性的提高，在果蝇可延长其寿命。有资料显示，在衰老过程中，eEF - 1α 的活性下降。因此，可能衰老过程中伴随着翻译忠实性的下降。另外，eEF - 1α 的大量表达还可以提高细胞的生长能力；一些引起细胞生长的因素也可引起细胞内 eEF - 1α 的大量表达，如癌基因 v-*fos* 可诱导 eEF - 1α 的表达。

此外，还发现 eEF - 1α 与细胞内其他成分如肌动蛋白等结合的现象，其意义不明。

(2) 延伸因子 2　延伸因子 2 是单体蛋白质分子。能与 GTP 结合，GTP 为其发挥功能所必需的。GTP 结合位点位于 eEF - 2 分子的 N 端区域。eEF - 2 具有 GTP 酶活性，该活性由 60 S 亚基的核糖体某一区域在移位后激活。对 GDP 的亲合性比 GTP 高出很多。eEF - 2 对核糖体的亲合性有两种状态，高亲合性出现在移位前，低亲合性出现在移位后。eEF - 2·GTP 和 eEF - 2·GDP 间的转变不需其他的因子参与。

eEF - 2 有二种共价修饰方式。His - 714 在多种酶的催化下产生修饰作用，这种修饰不是其功能所必需的，但为第二种修饰的基础，即只有发生这种修饰后，eEF - 2 才能进行 ADP -核糖基化，核糖基化的 eEF - 2 完全失去了 GTP 酶活性，不能与核糖体结合，失去了结合 GTP 及 GDP 的能力。第三种修饰方式是磷酸化。磷酸化的位点在 Thr56 和 Thr58，二者均位于 GTP 结合的区域。磷酸化的结果是使 eEF - 2 失活。

磷酸化是在 eEF - 2 激酶的催化下完成的。该酶的催化亚基为 100 kDa，与钙调蛋白

结合。催化亚基的活性完全依赖于钙离子和钙调蛋白。该酶在自身催化或 cAMP 依赖性蛋白激酶催化磷酸化后，对钙离子和依赖程度降低。因而，钙离子和 cAMP 可以抑制延伸。

(3) 延伸因子 3　　延伸因子 3 只在酿酒酵母(*Saccharomyces cerevisiae*)、白色念珠菌(*Candida albicans*)和裂殖酵母(*Schizosaccharomyces pombe*)中发现。由一条多肽链组成，分子量为 120～125 kDa，具有结合和水解 ATP 和 GTP 的能力，二种活性位于同一活性位点。在一级结构中有一重复的核苷酸结合区域，与其他 ATP 结合蛋白相似。在结构上与原核生物核糖体 S5 蛋白和一些氨基酰- tRNA 合成酶有相似之处。

eEF－3 的功能尚不十分清楚。它对每一轮延伸都是必需的，可能起到保证翻译准确性的作用，类似于原核生物核糖体的 S5 蛋白。有人提出酵母核糖体上存在 I 位点，eEF－1α·GTP·aa-tRNA 首先进入 I 位，而后在 eEF－3 的作用下进入 A 位。为何高等真核生物没有 eEF－3，目前尚不清楚。

2. 延伸循环

延伸循环可分为三个阶段：进位、肽键形成和移位。

(1) 进位　　进位是指 aa-tRNA 进入 A 位。aa-tRNA 是以 eEF－1α·GTP·aa-tRNA 复合物的形式进入 A 位的，并需要 GTP 的水解。在这一过程中，如何保证正确 aa-tRNA 的进入即翻译的准确性是一关键问题。目前有两种模型，一种认为对 aa-tRNA 经过了二步选择，第一步是在三元复合物进入 A 位时；第二步是在 GTP 水解后和肽键形成前。另一种是三位点模型，认为除 A 和 P 位外，还存在一个 aa-tRNA 结合位点。在原核生物第三个位点(E 位点)已被证实，而真核生物资料还不充分。已发现兔肝 80 S 核糖体有三个结合 aa-tRNA 的位点。

(2) 肽键形成　　aa-tRNA 进位后，立即形成肽键。肽键形成是由核糖体大亚基的肽酰转移酶中心的催化下完成的。A 位与 P 位间的距离要比两种 aa-tRNA 发生反应所需距离大，因此推测在形成肽键时核糖体的空间构象要发生扭曲。氨基酰- tRNA 的 α 氨基对肽酰- tRNA 羧基发动亲核攻击，从而形成肽键。

(3) 移位　　移位是指肽键形成后，核糖体沿 mRNA 向 3′方向移动一个密码子的距离。对真核生物的移位过程了解还不十分清楚，可能与原核生物相似。移位需 eEF－2 参与，移位后可使 eEF－2 的构象发生变化，从而水解 GTP，对核糖体亲合力下降并游离下来。但在核糖体缺乏 eEF　1α·GTP·aa-tRNA 时，eEF－2 可能依然维持与核糖体结合的状态，当三元复合物与核糖体结合后，eEF－2 才能离开。有些资料表明，eEF－2 并不是延伸所必需的，它只起到加速的作用。

三、合成终止

合成终止是在终止因子的作用下肽链停止延伸及核糖体与 mRNA 分离的过程。尽管在 70 年代发现了释放因子和终止密码，但至今还未对其精确机制有较详细的了解。

1. 真核生物的终止因子

真核生物只有一种释放因子(eukaryotic releasing factor,eRF),能识别三种终止密码子 UAA、UAG 和 UGA。最初发现的是兔的 eRF,分子量为 56.5 kDa。试验表明 eRF 不能终止细菌核糖体的肽链延伸反应,而且原核生物有三种释放因子,说明真核生物和原核生物在终止的机制上存在差异。应用兔 eRF 的单克隆抗体筛选表达文库,已得到了编码该蛋白的 cDNA,发现在序列上与哺乳动物的色氨酰- tRNA 合成酶有大于 90% 的同源性。但二者在热稳定性和磷酸纤维素层析特性上都有很大区别,WRS 也不具有 eRF 的活性,说明二者是不同的二种蛋白质。但为何二者有如此高的同源性,尚待探讨。

在酿酒酵母中,已发现了 *SUP45/SAL4* 的基因产物 Sup45p 有终止因子的活性,但可能只是一种类似于原核生物 RF-3 的辅助因子。

线粒体有自己独特的终止因子(mitochondrial releasing factor,mRF)。在酵母的线粒体中发现了 mRF-1,与大肠杆菌的 RF-1 和 RF-2 分别有 38% 和 30% 的同源性。它识别 UAG 和 UGA 的能力还不十分清楚。大鼠线粒体的 RF 在识别终止密码子的特性上和细菌的 RF-1 相似。

2. 终止的机制

当核糖体移动到终止密码子(UAA、UAG 和 UGA)处时,由于没有氨基酰-tRNA 能对其识别,核糖体将在此处暂停,eRF 因子可以识别终止密码子,并与核糖体形成复合物,引起肽链的释放,在核糖体释放因子(ribosome releasing factor,RRF)共同作用下,核糖体与 mRNA 解离,肽链延伸终止。

(1) 终止密码子　无论原核生物还是真核生物,终止密码子均为 UAA、UAG 和 UGA(线粒体有例外)。但通过分析大量的终止子发现,终止密码子的使用频率是不同的,且密码子的下游的第一个核苷酸并不是随机的。因而有人提出终止密码子应由 4 个核苷酸组成而不是我们通常认为的三个核苷酸。体外试验证明,真核生物的 RF 需要识别 4 个核苷酸才能终止翻译。资料表明有些真核生物的终止密码子的下游核苷酸倾向于 G。但上述观点还缺乏体内试验的证据。另外,新生肽链的氨基酸种类对终止的效果也有影响。

(2) 终止密码子的识别　翻译的终止是通过 RF 与终止密码子的相互作用完成的。RF 可识别第一个进入到 A 位的终止密码子,并导致肽酰-tRNA 的水解。至于 RF 的结合位点有以下三种模型:一是认为 RF 直接与处于 A 位的终止密码子结合;第二种认为某种 rRNA 通过碱基配对与 A 位的终止密码子结合,RF 识别这种双链 RNA 结构;第三种认为虽然存在 rRNA 与 mRNA 间的相互作用,但 RF 不是识别这种双链 RNA 的结构,而是由于 rRNA 与 mRNA 的相互作用引起了 rRNA 和核糖体的构象发生变化,使 RF 可以识别终止密码子以外的某一区域,如肽酰转移酶。

在体外实验中已证实,RF 与终止密码子间特异的相互作用类似于碱基配对,也就是说类似于延伸过程中 tRNA 的反密码子与 mRNA 的密码子间的相互作用。在原核生物中,已发现 30 S 亚基的 16 SrRNA 参与了 RF 识别终止密码子的过程。在真核生物,这方面的报道还很少。

(3) 翻译的终止　在终止因子识别终止密码子后，肽酰- tRNA 水解，新生肽链释放，核糖体与 mRNA 解离，60 S 与 40 S 亚基解离。解离后的大小亚基可参与下一次循环。有关具体的机制，真核生物了解的还不多。

第六节　蛋白质生物合成初始产物的后加工

由核糖体释的新生肽链并不是一个完整的、有生物学功能的蛋白质分子，必须经过后加工，才具有生物学活性，包括形成高级结构、与其他亚基缔合及其他的共价修饰。

一、肽链中氨基酸残基的化学修饰

对新生肽链中某些氨基酸残基进行共价修饰，是翻译后处理的重要内容。反应的主要类型有以下几种：

(1)乙酰化　主要发生在 N 末端的 α 氨基和赖氨酸的 ε 氨基上；

(2)甲基化　发生在 α 氨基、ε 氨基、精氨酸的胍基和 C 末端的 α 羧基和侧链的羧基上；

(3)磷酸化　主要发生在丝氨酸的羟基和苏氨酸及酪氨酸的羟基上；

(4)泛酸化　发生在 α 氨基和 ε 氨基上；

(5)转氨基酸作用　主要发生在 N 末端的 α 氨基上；

(6)多聚 ADP 核糖基化　主要发生在精氨酸的胍基上；

(7)糖基化　可以通过 N 糖苷键连接于天冬氨酰的酰胺基上，也可以通过 O 糖苷键连接于丝氨酸和苏氨酸的羟基上以及羟赖氨酸和羟脯氨酸的羟基上，也可以通过 S 糖苷键连接于半胱氨酸的巯基上。

在每一种类型的共价修饰过程中，可以由不同的酶催化，如 α 氨基和羧基的甲基化是在不同的酶催化下完成的。每种共价修饰都具有重要的生理功能。

二、肽链 N 端甲硫氨酸或甲酰甲硫氨酸的除去

在原核生物中，蛋白质起始合成的第一个氨基酸为甲酰甲硫氨酸；真核生物为甲硫氨酸。而成熟的蛋白质 N 末端大部分不是甲硫氨酸，故必须切去 N 端的一个或几个氨基酸。该氨基酸的甲酰基可由脱甲酰酶催化下而被去除。在多数情况下，当肽链的 N 端游离出核糖体后，立即进行脱甲酰化。N 端的甲硫氨酸的去除也可在合成起始后不久发生，但这一过程受肽链折叠的影响。

三、信号肽的切除

无论原核生物还是真核生物，蛋白质除游离于胞浆内发挥作用外，还有一部分要分泌到细胞外和定位于膜系统中起作用。如革兰氏阴性菌的细胞膜具有两层膜结构，内膜含有与能量代谢和物质转运有关的蛋白质；外膜有促进离子和营养物质进入细胞的蛋白质，

膜周质中含有水解酶及其他蛋白质。真核生物细胞更加复杂，细胞内有各种不同的细胞器，每种细胞器又有不同的膜结构。蛋白质在细胞内的定位更加复杂。蛋白质不但要决定是否越膜，还要决定要越膜的种类。在越膜过程中，有时要在翻译的同时发生一些处理过程(cotranslational processing)，而翻译后发生的共价修饰称为翻译后加工(posttranslational processing)。

在绝大多数细菌的越膜蛋白中，N端都有一长为15～30个氨基酸的信号肽。信号肽的前半段富含精氨酸和赖氨酸，带正电荷；后半段以疏水性氨基酸为主，约有15～20个氨基酸。该段氨基酸能形成α螺旋结构，紧接信号肽的序列也能形成α螺旋。两段α螺旋以反向平行形成一个发夹结构，很容易进入脂双层。继续合成的肽链能穿过内膜。如果是内膜蛋白，则会有一段或几段疏水区域结合于内膜中。当组成发夹结构的两段α螺旋间的部分达内膜外表面时，位于内膜外表面的信号肽酶(signal peptidase)就切除信号肽。翻译越膜蛋白的核糖体是与内膜结合的，这更便于越膜过程。

真核生物细胞中合成的蛋白质如何到达不同的部位取决于蛋白质是否具有转运信号。如果没有转运信号，蛋白质就保留在细胞质中。不同的细胞定位需要不同的转运信号。细胞表面蛋白、分泌蛋白和溶酶体蛋白(统称为分泌蛋白)具有与原核生物相似的信号肽，位于N端，长约15～30个氨基酸残基，大部分为疏水性氨基酸，接近N端有几个正电荷的精氨酸和赖氨酸。真核生物的信号肽也能形成α螺旋的发夹结构，在信号肽识别颗粒(signal recognition particle，SRP)的协助下插入到糙面内质网的膜中。SRP是一个RNA和蛋白质组成的复合物，由7 SL RNA和6种蛋白质组成。其中，7 SL RNA由300个核苷酸组成。SRP由两个结构域，一个是信号肽识别结构域，由7 SL RNA中的100至250个核苷酸部分与19、54、68和72 kDa蛋白质构成，另一个由7 SL RNA的其他部分与9和14 kDa蛋白质组成。SRP能与糙面内质网上的锚定蛋白(docking protein)结合，并引导新合成的肽链进入糙面内质网。最后引导分泌蛋白质的越膜。合成的蛋白质而后进入高尔基复合体接受修饰，或固定于膜中，或形成分泌小泡，分泌到细胞外。目前对核内、线粒体和叶绿体的蛋白质转运机制还不清楚。

四、肽链的折叠

肽链的折叠在肽链合成没有结束时就已经开始。核糖体可保护30至40个氨基酸残基长的肽链，当肽链从核糖体中露出后，便开始折叠，三级结构的形成几乎和肽链合成的终止同时完成。例如，大肠杆菌β半乳糖苷酶的抗体可识别酶的三级结构，能与合成该酶的多核糖体结合。该酶的活性形式为四聚体，当新生的肽链还未由核糖体释放时，就能与游离的酶分子形成活性形式。这说明高级结构的形成在合成终止前就开始了。由此可见，蛋白质的折叠是从N端开始的。

五、切除前体中功能不必需肽段

在蛋白质的前体分子中，有一些肽段是功能所不需要的，在成熟的分子中不存在。肽段的切除是在专一性的蛋白水解酶的作用下完成的。如前胰岛素原的加工过程中就去除

了分子内部的连接肽(C肽)。多种多肽激素和酶的前体大都要经过这一加工过程。

六、二硫键的形成

在mRNA分子中,没有胱氨酸的密码子,而不少蛋白质分子中含有胱氨酸二硫键,有的还有多个。且二硫键是蛋白质的功能基团。二硫键是通过两个半胱氨酸的巯基氧化形成的,有的在切除肽段前就已形成。

七、多肽链N端和C端的修饰

在少数情况下,合成的多肽一或两端存在修饰氨基酸,如微管蛋白的α链在酶的作用下,C端能被酪氨酸修饰,且不需要tRNA。还发现一种能从活化的tRNA上将氨基酸残基转移到成熟的蛋白质N端上。哺乳动物中也有类似的现象,意义不详。

在病毒和细菌中,有些蛋白质的N端氨基被乙酰化。而在真核生物中,细胞中有半数以上的蛋白质N端被乙酰化。乙酰化受N乙酰转移酶催化,该酶对N端氨基酸有选择性,能被修饰的有:甘氨酸、丙氨酸、丝氨酸、甲硫氨酸和天冬氨酸。这种作用的生物学功能不详,可能具有调解蛋白质稳定性的作用。多数多肽的C端被酰胺化,特别是多肽激素,如催产素、加压素、促胃液素、缩胆囊素和分泌素。酰胺化能保护多肽免受外切酶的水解。另外,N端还有葡萄糖胺和脂肪酸基团的修饰等。

(张永亮)

参考文献

[1] Losick R and Chamberlin M. New York: Cold Spring Harbor Laboratory. 1976
[2] Doi R H et al. Microbil Rev. 1986 50:227~243
[3] Losick R et al. Ann Rev Genet. 1986. 20:625~669
[4] Helmann J D et al. Ann Rev Biochem. 1988. 57:839~872
[5] Rice G A et al. Proc Nat Acad Sci USA. 1991. 88:4245~4249
[6] Haldenwang W G et al. Cell. 1981. 23:615~624
[7] Grossman A D et al. Cell. 1984. 38:383 -390
[8] Haldenwang W G. Microbiol Rev. 1995. 59:1~30
[9] Yong R A. Ann Rev Biochem. 1991. 60:189~715
[10] Sakonju S et al. Cell. 1980. 19:3~15
[11] Sakonju S et al. Cell. 1980. 19:27~35
[12] Kassavatis G A. Cell. 1990 60:235~245
[13] Buratowski S. Cell. 1989. 56:549~561
[14] Zenke M et al. EMBO J. 1989. 5:387~397
[15] Mueller-Storm H P. Cell. 1989. 58:767~777
[16] Stephen T Smale. Biochimica et Biophysica Acta. 1997. 1351:73~88
[17] Steve Busby. Cell. 1994. 79:743~746
[18] Adhya S et al. Ann Rev Biochem. 1978. 47:967~996
[19] Platt T. Ann Rev Biochem. 1986. 55:339~372

[20] Roberts T W. Nature. 1969. 224:1168~1174
[21] Greenblatt J et al. Cell. 1981. 24:421~428
[22] Yager R D et al. 1987. Ed. F. C. Neidhardt, Washington DC, 1987. 1241~1275
[23] Das A. Ann Rev Biochem. 1993. 62:893~930
[24] Geiduschek E P et al. Ann Rev Biochem 1988. 57:873~914
[25] Zawel L et al Prog Nucleic Acids Res Mol Biol. 1993 44:67~108
[26] Burley S K et al. Ann Rev Biochem. 1996. 65:769~800
[27] Friedberg E. Ann Rev Biochem. 1996. 65:14~42
[28] Kim Y et al. Nature. 1993. 365:512~520
[29] Goodrich J A et al. Cell. 1993. 77:145~156
[30] Martinez E et al. EMBO J. 1994. 13:3115~3126
[31] Nikolov D B et al. Nature. 1995. 377:199~228
[32] Verrijzer L P et al. Cell. 1995. 81:1115~1125
[33] Leigh Zawel Ann Rev Biochem. 1995. 64:533~561
[34] Stanley N C. Cell. 1995. 80:829~832
[35] Banerjee A K. Microbiol Rev. 1980. 44:175~205
[36] Jackson R J et al. Cell 1990. 62:15~24
[37] Ross J. Microbiol Rev. 1995. 59:423~450
[38] Gegenheimer P et al. Microbiol Rev. 1981. 45:502~541
[39] Kole R et al in the enzymes, ed. Boyer P D. 1982. 15:570~483
[40] Sharp P A. Science. 1987. 235:766~771
[41] Guthrie C et al. Ann Rev Genet. 1988. 22:387~419
[42] Grthrie C. Science. 1991. 253:157~163
[43] Nilson T. Ann Rev Microbiol. 1993. 47:413~440
[44] Weiner A. Cell. 1993. 72:161~164
[45] Sharp P A. Cell. 1994. 77:805~815
[46] Kramer A Ann Rev Biochem. 1996 65:367~410
[47] Zhang Y et al. Cell. 1980. 46:827~835
[48] Zhang Y et al Proc Nat Acad Sci USA ,989 86:2752 2756
[49] Reed R et al. Cell. 1988. 53:949~961
[50] Sontheimer E J et al. Science. 1993. 262:1989~1996
[51] Hannon G J et al. Cell. 1980. 61:1247~1255
[52] Bruzik J P et al. Proc Nat Acad Sci USA. 1995. 92:7056~7059
[53] Cech T R. Ing Rev Cytol. 1985. 93:3~22
[54] Cech T R. Science. 1987. 236:1532~1539
[55] Cech T R. Ann Rev Biochem. 1990. 59:543~568
[56] Lambowitz A M et al. Ann Rev Biochem. 1993. 62:587~622
[57] Michel F et al. Ann Rev Biochem. 1995. 64:435~461
[58] Gert J A. Biochimica et Biophysica Acta. 1996. 1307:39~54
[59] Hiten D M et al. Ann Rev Genet. 1994. 28:1~26

第七章　原核生物基因表达调控

第一节　概　　述

机体能在基因表达过程的任何阶段进行调控，如调控可在转录阶段、转录后加工阶段和翻译阶段进行。

转录的调控主要发生在起始阶段，这样可避免浪费能量合成不必要的转录产物。通常不在转录延伸阶段进行调控，但可在终止阶段进行调控，终止可以防止越过终止子而进行下一个基因的转录。

RNA 的初级转录产物本身是一个受调控的靶分子，转录物作为一个整体其有效性可以受到调控，例如，它的稳定性可以决定它是否保存下来用于翻译。此外，初级转录产物转变为成熟分子的加工能力可决定最后 mRNA 分子的组成和功能。在真核细胞中，还可对 RNA 从核到胞浆中的转运进行调控。但是在细菌中，mRNA 只要一合成，就可用于翻译。

翻译也像转录一样，在起始阶段和终止阶段进行调控。DNA 转录的起始和 RNA 翻译的起始路线也很相似。

在原核生物和真核生物最常见的调控是转录过程的调控。因此本章先讨论转录调控，然后，再介绍翻译水平的调控。

为了便于理解，在介绍具体的调控过程之前，先介绍一些基本概念。

1. 顺式作用元件和反式作用因子

基因活性的调控主要通过反式作用因子（通常是蛋白质）与顺式作用元件（通常在 DNA 上）相互作用而实现。基因是编码可扩散产物的 DNA 序列，基因所编码的产物可以是蛋白质（大多数基因都编码蛋白质），也可以是 RNA（tRNA 和 rRNA）。其最重要的特点是基因产物将从合成的场所扩散到其发挥作用的其他场所。游离基因产物扩散至其目标场所的过程称为反式作用（*trans*-acting）。因此反式作用因子（*trans*-acting factor）的编码基因与其识别或结合的靶核苷酸序列不在同一个 DNA 分子上。

顺式作用（*cis*-acting）的概念用于任一不转变为任何其他形式的 DNA 序列，它只在原位发挥 DNA 序列的作用，它仅影响与其在物理上相连的 DNA。有时顺式调节序列最终发挥作用的分子不是 DNA，而是 RNA。因此，顺式作用元件（*cis*-acting element）是指对基因表达有调节活性的 DNA 序列，其活性只影响与其自身同处在一个 DNA 分子上的基因；同时，这种 DNA 序列通常不编码蛋白质，多位于基因旁侧或内含子中。

2. 结构基因和调节基因

为了区分调控过程中的调控成分和其调控的基因，有时用结构基因和调节基因的概念。结构基因（structural gene）是编码蛋白质或 RNA 的任何基因。结构基因编码着大量

功能各异的蛋白质，所编码的蛋白质有组成细胞和组织基本成分的结构蛋白、催化活性的酶和调节蛋白等。调节基因(regulator gene)是参与其他基因表达调控的 RNA 和蛋白质的编码基因。

调节基因编码的调节物通过与 DNA 上的特定位点结合控制转录是调控的关键。调节物与 DNA 特定位点的相互作用能以正调控的方式(启动或增强基因表达活性)调节靶基因，也能以负调控的方式(关闭或降低基因表达活性)调节靶基因。DNA 位点通常位于受调节基因的上游，但有时也有例外。

3. 启动子和终止子

位于转录单位开始和结束位置上的序列为启动子(promoter)和终止子(terminator)，两者都是典型的顺式作用位点。启动子位于基因转录起始点的上游，负责基因转录的起始。终止子能终止基因的转录。启动子和终止子是能受同一类反式作用因子识别的顺式作用元件，这一类反式作用因子就是 RNA 聚合酶。当然，两个位点也能各自结合一些特定的其他因子。

4. 操纵基因和阻抑蛋白

细菌的传统控制机制是负调控：即阻抑蛋白(repressor)阻止基因表达。在缺省的情况下(即无阻抑蛋白时)，RNA 聚合酶能够识别受调节基因的启动子，使这种基因得到表达。另一个叫做操纵基因(operator)的顺式作用位点与启动子相邻，操纵基因是阻抑蛋白的靶位点。当阻抑蛋白与操纵基因结合时，就会阻止 RNA 聚合酶启动转录，基因的表达因而被关闭。

5. 转录因子

除了上述负调控，基因转录的调控也有正调控。在细菌中，正调控与负调控使用的频率也许大致相等，但在真核生物中，最常见的调控方式是正调控。参与正调控的反式作用因子就是转录因子(transcription factor)。转录因子是转录起始过程中，RNA 聚合酶所需的辅助因子。典型真核基因的缺省状态(即无转录因子)是基因处于无活性状态，RNA 聚合酶自身不能在启动子上起始转录。反式作用因子在启动子附近有作用位点，其中的一些或所有因子的结合使 RNA 聚合酶能起始转录。

正调控和负调控的共同点是调控蛋白都是反式作用因子，它通常识别位于基因上游的顺式作用元件。不同类型的调节蛋白识别不同的序列。尽管调节蛋白结合 DNA 的实际长度较长，但它仅识别 DNA 上很短的序列，一般小于 10 bp。细菌启动子就是一个例子，尽管 RNA 聚合酶在转录起始时，覆盖大于 70 bp 的 DNA，但其识别的关键序列是位于 −35 和 −10 中心处 6 个碱基的序列。

原核生物和真核生物的基因组织差异很大，细菌的结构基因一般成簇排列，而真核生物基因则是独立存在。成簇排列的结构基因能受单一启动子共同控制，结果是整套基因或者都表达或都不表达。本章主要讨论细菌基因表达的控制模型，真核生物基因表达的调控在第八章介绍。

第二节 操 纵 子[1,2~6]

原核生物的调控可发生在转录和翻译等不同阶段，但以转录水平调控为主。大肠杆菌的乳糖操纵子是目前转录水平最清楚的调控，转录水平调控的许多一般规律都源于对乳糖操纵子的研究。

在细菌中编码功能相关的结构蛋白的基因常成簇排列在一起。例如，同一个代谢途径的酶的基因就常成簇排列。此外，除了参与代谢途径的酶以外，其他与该途径有关的蛋白质的基因也可能包括在此控制单元内。例如，负责转运小分子底物进入细胞内的蛋白质的基因有时就包括在这样的控制单元内。这样的控制单元就是操纵子(operon)。操纵子就是由操纵基因以及相邻的若干结构基因所组成的功能单位，其中结构基因转录受操纵基因(operator)所控制。操纵子中的全部结构基因从同一个启动子开始转录成单个mRNA分子。操纵基因则是操纵子的前端部分，它能和阻抑蛋白结合，控制有关操纵子的转录。因此操纵基因是顺式控制元件，阻抑蛋白是由调节基因表达出的蛋白质，是参与操纵子调节的反式作用因子。

一、乳糖操纵子的结构

参与β-半乳糖苷(如乳糖)的分解代谢的三个酶的基因就成簇排列，构成乳糖操纵子(lactose operon，*lac*)。图7-1列出了乳糖操纵子的结构基因、相关的顺式作用调控元件及反式作用调节基因的组织排列。

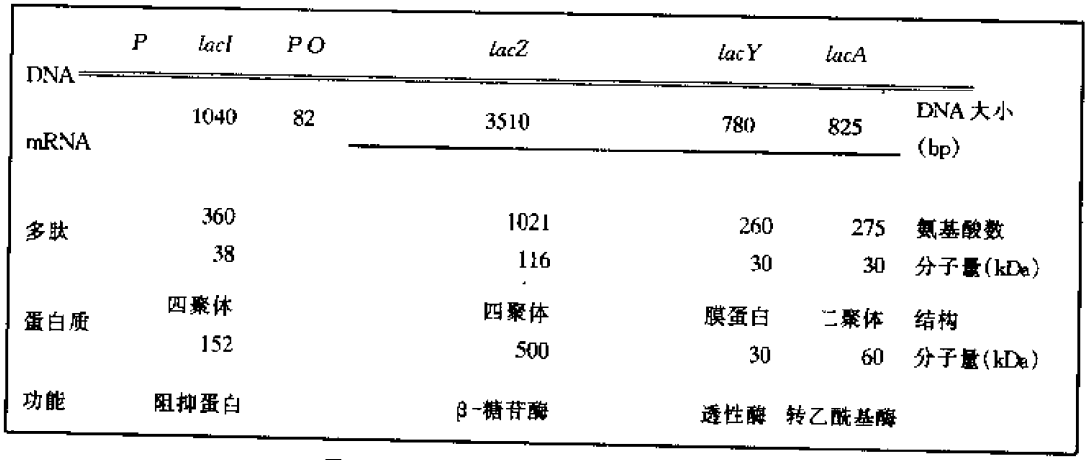


图7-1 乳糖操纵子约6000 bp DNA的结构

在左侧的*lacI*基因有自己的启动子(*P*)和终止子。*lacI*区的末端与启动子(*P*)相邻。操纵基因(*O*)含*lacZ*基因的前26 bp序列。

三个结构基因分别编码三种蛋白质。*lacZ*编码β-半乳糖苷酶。β-半乳糖苷酶是分子量约500 kDa的四聚体，能催化β-半乳糖苷水解，它能将乳糖水解为葡萄糖和半乳糖。*lacY*编码β-半乳糖苷透性酶，β-半乳糖苷透性酶是分子量30 kDa的膜结合蛋白，是β-半乳糖苷转运系统的组成成分，参与β-半乳糖苷进入细胞内的转运。*lacA*编码β-半乳糖苷转乙酰基酶，β-半乳糖苷转乙酰基酶催化将乙酰基团从乙酰辅酶A转移到β-半乳

糖苷上。

lacZ 或 *lacY* 突变能产生不能利用乳糖的 *lac* 基因型。*lacZ* 突变导致酶失活，因而不能分解利用乳糖。*lacY* 突变菌不能从培养基中吸收乳糖。令人不解的事，仍没有分离到 *lacA* 缺陷的细菌。细菌在含不能被分解的某些 β-半乳糖苷类似物的培养基中培养时，很可能乙酰化反应发挥作用，因为这种修饰可引起脱毒和外排。

结构基因突变只引起其编码的特定蛋白失活，而调节基因突变则影响其控制的所有结构基因的表达。*lacZYA* 基因的转录受 *lacI* 基因指导合成的调控蛋白的控制。*lacI* 基因正好与结构基因相邻，但它不与结构基因属同一转录单位，它有自己独立的转录单位，含自己的启动子和终止子。*lacI* 编码可扩散的产物，理论上，它不必位于结构基因附近，*lacI* 基因移到其他任何地方都能很好地发挥作用，因此它属于反式作用调节因子。

乳糖操纵子的控制属负调控，也就是说调节蛋白关闭其转录。*lacI* 的表达产物称为乳糖操纵子阻抑蛋白（*lac* repressor），它的功能是阻止结构基因的表达。乳糖操纵子阻抑蛋白是由 4 个同样亚基组成的四聚体，每个亚基分子量为 38 kDa。一个大肠杆菌细胞中有大约 10 个阻抑蛋白分子。调节基因转录成单顺反子 mRNA，转录的速度似乎只受其启动子与 RNA 聚合酶的亲和力大小的控制。

阻抑蛋白通过与 *lacZYA* 基因簇开始处的操纵基因（用 O_{lac} 表示）结合发挥功能。操纵基因位于启动子（P_{lac}）和结构基因（*lacZYA*）之间。阻抑蛋白与操纵基因结合时，能阻止 RNA 聚合物在启动子上的转录起始。*lac* 操纵基因是 mRNA 起始点上游 −5 bp 至转录单位内 +21 bp 的序列，因此它与启动子的右末端重叠。

二、小分子诱导物对阻抑蛋白活性的影响

细菌需要快速应答环境的变化，营养供给随时都可能发生变化，其生存取决于从分解利用一种物质转换为分解利用另一种物质的能力，因此对细菌来说，灵活性就特别珍贵。当然，经济性也相当重要，因为如果一个细菌以放任能量消耗的方式去适应环境则有很大的弊端。在某种物质并不存在时，产生参与该物质代谢的各种酶，代价实在太昂贵。细菌的做法是当某种物质不存在时，并不合成该物质代谢通路中的酶，但随时准备这种物质一出现就可合成所需的酶。应答某种特定物质出现而合成特定酶的过程称为诱导（induction），这种调控方式在细菌中很普遍，在单细胞真核生物（如酵母中）也存在。大肠杆菌乳糖操纵子是这种控制机制的最好范例。

当大肠杆菌在不含 β-半乳糖苷的环境中培养时，不需要 β-半乳糖苷酶，细胞内含的 β-半乳糖苷酶很少，每个细胞不到 5 个分子。加入适宜的底物后，细胞内很快出现酶活性，2～3 分钟内，一些酶就出现，且很快达每细胞 5000 个分子的水平。在适宜的条件下，β-糖苷酶的量可达细胞总可溶蛋白的 5%～10%。如果底物从培养基中去掉，酶的合成很快停止到诱导前的水平。

lac mRNA 极不稳定，其半衰期仅有约 3 分钟，这使得诱导能很快逆转。只要诱导物一除去，转录就停止，所有 *lac* mRNA 便在很短的时间内分解掉，细胞内的 mRNA 含量回到诱导前的基础水平。

这种快速应答环境营养成分变化的能力不仅表现在分解新底物方面，而且也用来关闭

突然出现在培养基中的化合物的内源性合成，例如，大肠杆菌通过色氨酸合成酶（tryptophan synthetase）催化合成色氨酸，但是，如果培养基中含有色氨酸，酶的合成立即停止。这种效应就是所谓的阻抑作用（repression），它使细菌避免采用自身资源进行不必要的合成。

诱导作用和阻抑作用都能对细菌的代谢能力进行调控。前者调节细菌使用某种物质的能力，后者调节合成特定代谢中间物的能力。每种调节的诱发剂都是小分子物质，或为酶底物，或为酶所催化合成的产物。如果某种小分子物质能够促使细菌产生酶将其自身分解，这种小分子物质就叫做诱导物（inducer）。如果某种小分子物质能够阻止细菌产生合成其自身的酶，这种小分子物质就叫做辅阻抑物（corepressor）。

诱导物或辅阻抑物的作用是高度特异的，仅有特定或密切相关的分子能发挥作用。但是这些小分子并不直接与其调控的目标酶分子相互作用而实现其调控功能。β-半乳糖苷酶的一些诱导物与天然诱导物相似，但不能被酶分解。例如异丙基硫代-β-D-糖苷（isopropylthio-β-D-galactoside，IPTG），尽管不能被β-半乳糖苷酶识别，但它是乳糖操纵子非常有效的诱导物。这种能诱导酶合成，但不能被酶分解的分子称为义务诱导物（gratuitous inducer）。义务诱导物由于能在细胞内保持原型不变，因此很有用，而真正的诱导物会被代谢分解掉，从而影响对系统的研究。用义务诱导物进行研究得到了一个重要结论：在诱导系统中，一定有一些与精确识别特异底物的酶分子不同的成分，这些成分能识别与底物相类似的分子，而酶不能。在乳糖操纵子中，这种应答诱导物的成分就是 *lacI* 编码的阻抑蛋白。它在应答环境变化时控制结构基因 *lacZYA* 转录的作用归纳于图 7-2。三个结构基因从 *lacZ* 上游的启动子开始转录成一条 mRNA。阻抑蛋白的状态决定启动子是开启还是关闭。缺乏诱导物时，由于阻抑蛋白具有结合操纵基因的活性，结构基因不被转录。而加入诱导物后，阻抑蛋白转变成非活性形式而离开操纵基因，因此转录从启动子开始，并通过结构基因，一直转录到位子 *lacA* 边上的终止子。

这种控制回路的关键在于阻抑蛋白的二重性：它既能阻止转录，又能识别小分子诱导物。阻抑蛋白上有二个结合位点，一个是操纵基因结合位点，另一个是诱导物结合位点。当诱导物与它的结合位点结合后，改变了阻抑蛋白质的构型，从而影响操纵基因结合位点的活性。

诱导实现协同调节（coordinate regulation），这种调节中所有受调节的基因作为整体表达或不表达。mRNA 从 5′端依次翻译，这解释了为什么诱导总是引起β-半乳糖苷酶、β-半乳糖苷酶透性酶和半乳糖苷转乙酰基酶依次出现，共用的 mRNA 的翻译解释了为什么在不同的诱导条件下三种酶的相对量总是保持不变。诱导是基因表达的开关，虽然不同的诱导物诱导效率不同，但这些因子只影响转录和翻译的绝对水平，并不影响三种基因的相互关系。

操纵子组成成分中有一个矛盾，乳糖操纵子含有编码分解该糖所需的半乳糖苷酶的结构基因（*lacZ*），也包括编码负责转运底物进入细胞的蛋白的基因（*lacY*）。但是如果操纵子处于阻抑状态，诱导物怎样进入细胞开始诱导过程呢？两个方面的机制保证了细胞内总有小量的蛋白，足以开始诱导，一是操纵子有一基础水平的表达，即使不被诱导时，操纵子也有少量表达（诱导水平的 0.1%）；另一个机制是一些诱导物通过其他吸收系统进入细胞内。

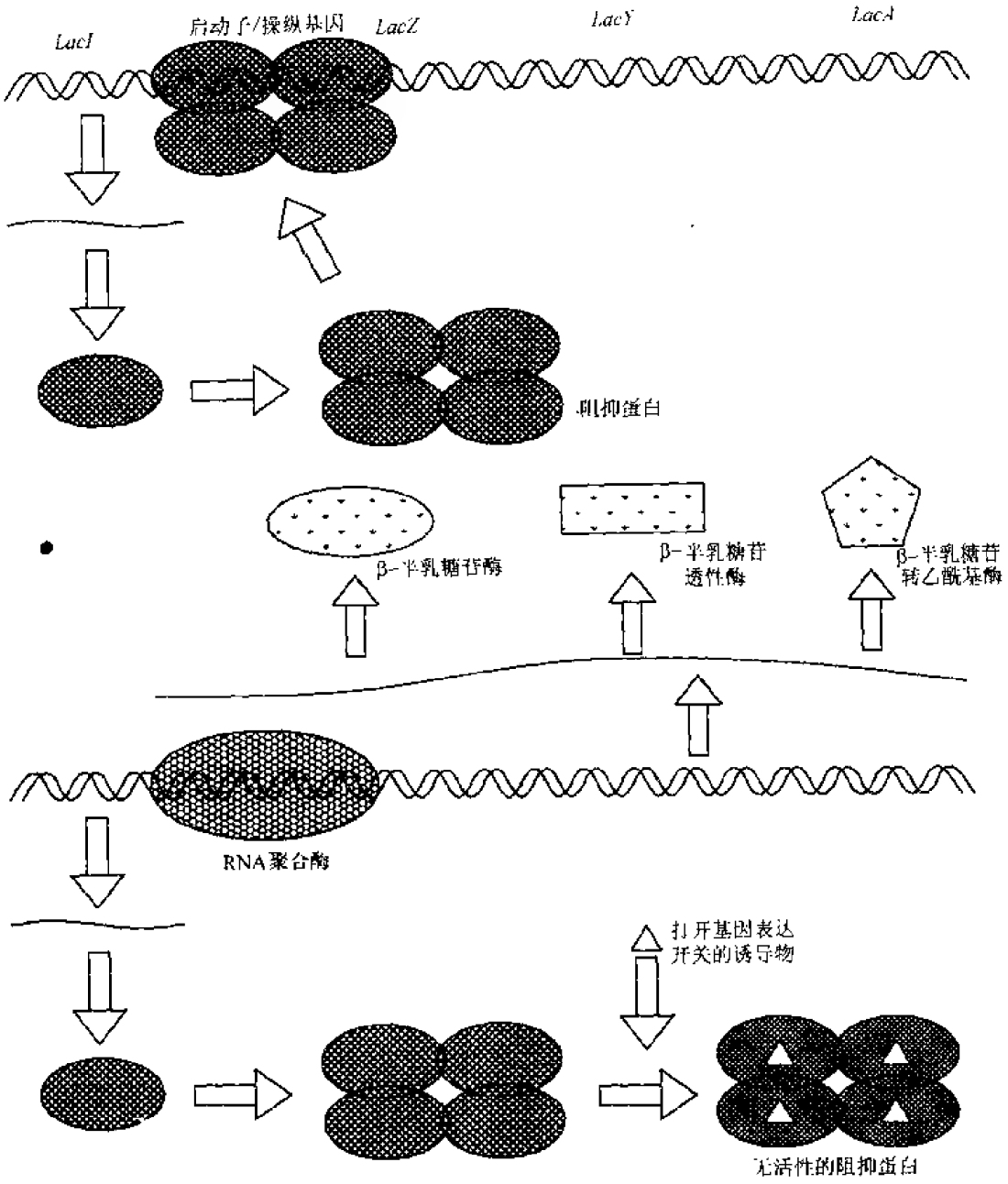


图 7-2 乳糖操纵子调控机制

阻抑蛋白通过与操纵基因结合，使乳糖操纵子通常处于非激活状态。加入诱导物，释放出阻抑蛋白，因此乳糖操纵子处于激活状态，RNA 聚合酶可以起始转录。

三、操纵子突变

调控通路中的突变可能消除操纵子的表达，也可能引起无论有无诱导物，操纵子都持续不断地表达。使操纵子在任何情况下，总不能表达的突变称为不可诱导(uninducible)的突变。而不受调控物影响的连续表达称为组成型基因表达(constitutive gene expres-

sion),操纵子的这种突变称为组成型(constitutive)突变。操纵子调控通路中的成分可通过并不位子结构基因内,并影响所有结构基因表达的突变来鉴定。经鉴定确定,启动子和操纵基因分别是 RNA 聚合酶和阻抑蛋白的靶元件,*lacI* 基因座编码阻抑蛋白。

1. 操纵基因突变与顺式显性

最早通过组成型突变鉴定了操纵基因,这种突变被称为 O^c,它为顺式作用元件提供了第一个证据。与发生 O^c 突变相邻的结构基因呈组成型表达,其机制是突变使操纵基因改变,不再与阻抑蛋白结合,因此阻抑蛋白不能阻止 RNA 聚合酶起始转录,于是操纵子持续不断地表达。

乳糖操纵子的操纵基因仅能控制与其相邻的 *lac* 基因。如果将第 2 个乳糖操纵子插入到独立的 DNA 分子中,引入细菌内,由于这个乳糖操纵子有自己的操纵基因,因此两个操纵基因互不干扰。如果一个操纵子带有野生型操纵基因,通常情况下它将处于阻抑状态,而第 2 个含 O^c 突变的操纵子则不被阻抑,仍按其特有的方式表达。这些特性表明操纵基因是一个典型的顺式作用位点,其功能取决于一些反式作用因子对其 DNA 序列的识别。操纵基因控制着邻近的基因,而对细胞内存在的其他等位基因无作用。此位点的突变,如 O^c 突变称为顺式显性(*cis*-dominance)突变。

顺式显性突变的概念适用于其功能是被识别而不是转换成可扩散性产物的任何 DNA 序列的突变。因此,顺式显性突变属于顺式作用位点突变,它不属于任何互补群(complementation group)。互补能力是产物可扩散的基因独有的性质。当两个顺式作用位点靠在一起时(例如启动子和操纵基因),我们就无法用互补测验(complementaion test)对突变分类,而只能根据其表型予以区分。

顺式显性是与其所控制序列相邻的任何顺式作用位点的一个特性。因此,如果一个控制位点属多顺反子 mRNA 的一部分,其突变也会呈现与其 DNA 突变同样的顺式显性模式,就像它们在 DNA 中所起的作用一样。总之,顺式显性的关键是,控制部位在位置上不能同基因分开。从遗传学的角度看,不论基因及其控制位点是共同位于 DNA 上还是共同位于 RNA 上都无关紧要。

2. 操纵子不可诱导型突变

操纵子不可诱导型突变在遗传学上可分成两类。每一类的显性关系都能用于确定基因座的性质。启动子突变是顺式作用。如果突变使得 RNA 聚合酶不能与 P_{lac} 结合,就会使操纵子不能转录而丧失功能。阻抑蛋白失去结合诱导物能力丧失的突变被称为 $lacI^s$。在这种突变中,阻抑蛋白被锁定为能识别操纵基因阻止转录的活性形式。由于阻抑蛋白上诱导物的结合位点缺失,加入诱导物也不起作用,所以不可能将阻抑蛋白转变成非活性形式。

3. 阻抑蛋白基因的 $lacI^-$ 突变

组成型转录也可由 $lacI^-$ 型突变引起,这种突变由 *lacI* 功能消失(包括基因缺失)形成。当阻抑蛋白失活或缺乏时,转录能在启动子处起始。也就是说,阻抑蛋白失去结合操纵基因能力的突变也呈现组成型基因型。将 $lacI^-$ 基因与野生型 $lacI^+$ 置于同一细胞中,

测定 $lacI^-$ 组成型突变基因和 $lacI^+$ 基因的关系能进一步证实这个结论。这种测定通过引入第二拷贝的操纵子来进行，第二拷贝的操纵子放在质粒(plasmid)中。质粒是只携带几个基因的能自主复制的 DNA 分子。在细胞中，如果有两个调节基因，一个调节基因是 $lacI^+$，另一个是 $lacI^-$，则恢复正常的调控。当诱导物去掉后，结构基因会再次受到阻抑。其机制是，第二个野生型调节基因引入后，正常的阻抑蛋白就会出现。因此一旦诱导物缺乏，操纵子就重新关闭。这表明阻抑蛋白是反式作用产物。

4. 各种突变的阻抑蛋白亚基的相互作用

活性阻抑蛋白的一个重要特点是它们形成多聚体，细胞内的阻抑蛋白亚基随机结合形成有活性的四聚体。当存在两个不同的 *lacI* 等位基因时，不同等位基因的亚基能相互缔合形成杂合四聚体，其性质不同于任何一种纯合四聚体。亚基间这种类型的相互作用是多亚基蛋白质的一个典型特征，称为等位基因间的互补(interallelic complementation)。

一些突变的阻抑蛋白间发生负互补作用(negative complementation)。$lacI^+$ 基因与 $lacI^{-d}$ 组合在一起时可见到负互补作用。单独的 $lacI^{-d}$ 突变导致阻抑蛋白不能与操纵基因结合，从而形成类似 $lacI^-$ 等位基因的组成型突变。因为 $lacI^{-d}$ 型突变使阻抑蛋白失活，所以它对野生型来说是隐性的。但是，$-d$ 这个标记指的是一种特殊的负调节类型，当它同野生型等位基因配合时，就表现出负调节性质，即呈现野生型的显性。这种突变称为反式显性(*trans*-dominant)，也称为显性负调控(dominant negative)。呈现显性的原因是因为 $lacI^{-d}$ 型等位基因产生了"坏"亚基，它不仅本身不能与操纵基因 DNA 结合，而且能参与形成阻抑蛋白四聚体，从而妨碍"好"亚基与操纵基因 DNA 结合。这表明阻抑蛋白完成阻抑作用，发挥作用的是四聚体，而不是单个亚基。

$lacI^{-d}$ 型突变位于阻抑蛋白亚基的 DNA 结合位点内，这解释了它们有能力阻止混合四聚体同操纵基因结合，结合位点数目的减少必定会大大降低特定的亲和性。从 *lacI* 的基因图谱(图 7-3)可见，$lacI^{-d}$ 突变成簇地位于基因的最左端。这就证明了阻抑蛋白靠

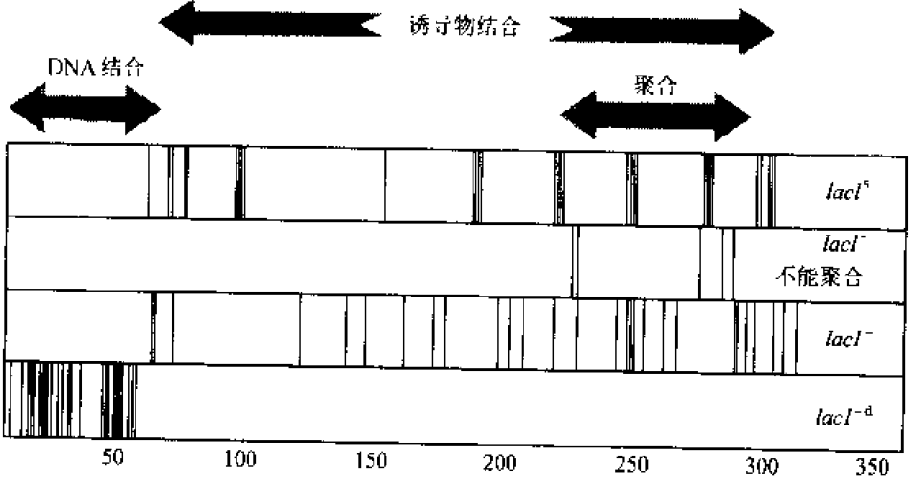


图 7-3 阻抑蛋白基因功能区突变图谱

$lacI^{-d}$ 型突变证明 DNA 结合区位于 N 末端区；不能形成四聚体的 *lacI* 突变位于 220～280 位基酸残基之间。另一种 $lacI^-$ 突变发生在整个基因中。$lacI^q$ 突变成簇地排列于氨基酸残基 62～300 间，突变点之间的间隔很有规律。

近 N 末端的区域是 DNA 结合部位。隐性的 *lacI*⁻型突变发生在分子中的其他任何部位，但它能间接影响 DNA 结合。

阻抑蛋白 N 末端区是特异性结合 DNA 的部位，也是“牢固结合”突变发生的部位，其作用已得到阐明。阻抑蛋白上这些部位的突变增强了其与操纵基因的亲和性，有时这种亲和力增加得太大，使得诱导物也不能使阻抑蛋白从操纵基因上释放下来，但这种情况并不多见。

不可诱导型 *lacI*ˢ 突变使阻抑蛋白对诱导物不起反应。这是因为阻抑蛋白失去了其结合诱导物的位点，或者因为它已不能将其作用传递给 DNA 结合位点。

四、阻抑蛋白作用机制

最初通过提纯能结合义务诱导物 IPTG 的成分的方法分离到阻抑蛋白。由于细胞内的阻抑蛋白量很少，为了得到足够的材料，有必要用启动子突变增加 *lacI* 的转录，并将 *lacI* 基因座放在一个细胞内有很多拷贝的 DNA 分子中，其结果是产量超过正常的 100～1000 倍。阻抑蛋白与含野生型乳糖操纵子序列的双链 DNA 结合，但它不与 O^c 突变的 DNA 结合。加入 IPTG 能从操纵基因 DNA 上释放阻抑蛋白。

1. 阻抑蛋白与特异 DNA 结合

那么阻抑蛋白怎样识别操纵基因 DNA 特异序列呢？操纵基因有细菌调节蛋白许多识别位点的共有特点就是具有对称性。如图 7－4 所示，以 +11 为对称轴，每边为两段各 6 bp 的对称序列 TGTGTG 和 AATTGT。靠近对称轴的一对反向重复序列在阻抑蛋白结合过程中起主要作用。甲基化试验表明，+3～+19 区段中的许多嘌呤碱基(A 和 G)可被阻抑蛋白保护而免受甲基化；相反，由于阻抑蛋白与操纵基因的结合而提高了少数嘌呤碱基对甲基化的敏感程度。这可能是由于在 DNA 阻抑蛋白的复合物中形成了一个疏水的口袋，从而加强了甲基化作用。而紫外线交联实验表明，位于 +1～+22 之间的绝大多数胸腺嘧啶在用 5－溴尿嘧啶取代后都能与阻抑蛋白形成交联，说明这些 T 都与阻抑

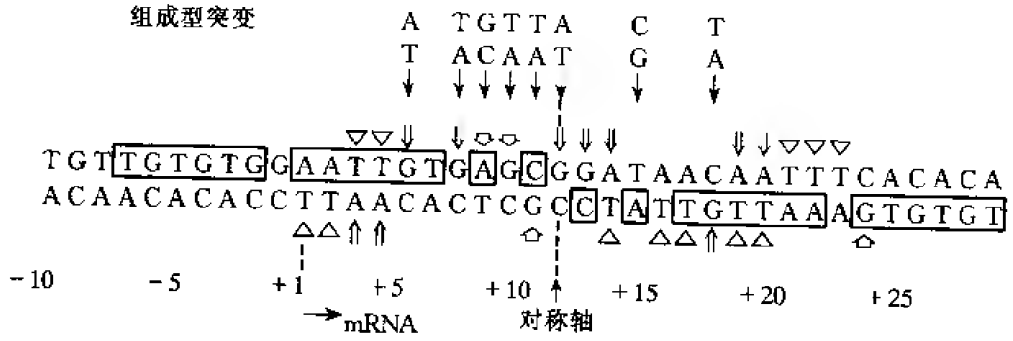


图 7－4 乳糖操纵子的对称序列

转录开始点定为 +1，框内的碱基序列示意对称序列。“▽”所指向的 T 在用 5－溴尿嘧啶取代后可与阻抑蛋白交联。“⇓”指向的嘌呤碱基可被阻抑蛋白保护而免遭甲基化。“⌓”指向的嘌呤碱基在阻抑蛋白与操纵基因结合后，其对甲基化的敏感程度提高。

蛋白紧密相连。可以用分析聚合酶与启动子相互作用的方法鉴定阻抑蛋白与操纵基因作用的位点。删除每一端的成分确定结合区的末端，对构成结合区的每个点进行突变确定必需的碱基。DNA 上与阻抑蛋白结合后受核酶保护的区位于对称区内，它是从 -5 至 +21 的 26 个碱基。组成型突变鉴定的区甚至更小，它位于从 +5 至 +17 的中心区内，有 8 个碱基，每个碱基被替换都会引起组成型突变。

DNA 序列的对称性表明蛋白质也应具有对称性。阻抑蛋白的结构确实具有对称性，它是同样亚基组成的四聚体，每个亚基都有同样的 DNA 结合位点。操纵基因的每个反向重复序列与操纵子一个单亚基以同样的方式接触。操纵基因的 +1 至 +6 与 +21 至 +16 完全相同也说明了这一点。但也应注意，操纵基因并不完全对称，左侧较右侧结合阻抑蛋白的能力更强。更强的操纵基因可以通过左侧的反向复制产生。

2. 诱导物对阻抑蛋白结合操纵基因的影响

在体外，不同诱导物使阻抑蛋白与操纵基因间亲和力降低的程度不同。其降低程度与诱导物在体内的诱导能力有关，这表明诱导作用是操纵基因与阻抑蛋白的亲和力降低而引起的。因此，当诱导物进入细胞时，诱导物降低操纵基因结合阻抑蛋白的亲和力。但是阻抑蛋白四聚体已紧紧地与操纵基因结合，诱导物怎样使其从操纵基因上释放下来呢？有两个模型说明阻抑蛋白怎样受诱导物作用从操纵基因释放出来（图 7-5）。

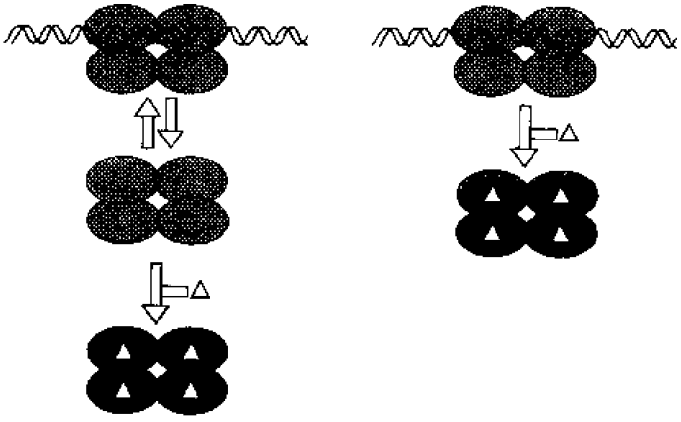


图 7-5 诱导物使阻抑蛋白释放的模型
左图模型中诱导物与游离的阻抑蛋白结合，打破阻抑蛋白与 DNA 结合的平衡。右图模型中诱导物的结合使阻抑蛋白直接从操纵基因中释放出来

一个模型是平衡模型（图 7-5 左侧）。该模型认为阻抑蛋白与 DNA 结合和释放是一个快速动态平衡过程，诱导物与游离的阻抑蛋白结合，使阻抑蛋白不能与 DNA 结合，这样便打破了这种平衡。但是阻抑蛋白从操纵基因上的解离速度太慢，用此模型难以解释。说明诱导物肯定与阻抑蛋白和操纵基因的复合物直接结合。如图 7-5 右侧所示，诱导物结合必须使阻抑蛋白产生变化，使其从操纵基因上释放下来。事实上，体外实验表明，加入 IPTG 能很快引起阻抑蛋白与操纵基因的复合物不稳定。

在甲基化保护/增强实验中，用较高浓度的蛋白质可对阻抑蛋白-IPTG 复合物与操纵基因的结合情况进行研究。大量的阻抑蛋白补偿了阻抑蛋白-IPTG 复合物对操纵基因过低的亲和力。实验结果表明阻抑蛋白-IPTG 复合物与 DNA 接触的方式与游离阻抑蛋白与 DNA 接触的方式确实相同。其他与操纵基因 DNA 有不同亲和力的突变的阻抑蛋白也呈现相同的实验结果。因此阻抑蛋白与 DNA 亲和力的变化是结合 DNA 的蛋白质整个构型改变的结果，而不是形成和打断一个或少数几个键造成的。

3. 阻抑蛋白结构与功能的关系

阻抑蛋白有几个结构域。DNA 结合结构域由 1～59 位残基形成，称为头段(headpiece)。它能用胰蛋白酶从阻抑蛋白的亚基中切下，切下头段后剩下的部分是核心段(core)。其详细的结构见图 7-6。阻抑蛋白单亚基的 N 端含两个 α 螺旋，两个螺旋之间被一个转折所分隔，这就是称为螺旋-转角-螺旋(helix-turn-helix，HTH)的 DNA 结合基序；两个 α-螺旋正好插到 DNA 的宽沟中，与特异 DNA 碱基接触。该 HTH 经一绞链区与蛋白质主体相连。在阻抑蛋白 DNA 结合形式中，绞链区形成如图 7-6 中所示的 α-螺旋；但是当阻抑蛋白不与 DNA 结合时，该区结构则杂乱无章。HTH 和绞链区一起构成头段。核心段由结构相似的两个区组成，每一区都有两个 α-螺旋夹着一个六链平行的 β 片层。诱导物结合在两区间的裂隙中。乳糖操纵了阻抑蛋白单亚基的 C 端有一 α-螺旋，其中含二个亮氨酸七聚体重复单位。四个亚基的螺旋连在一起，维持阻抑蛋白的四聚体结构。

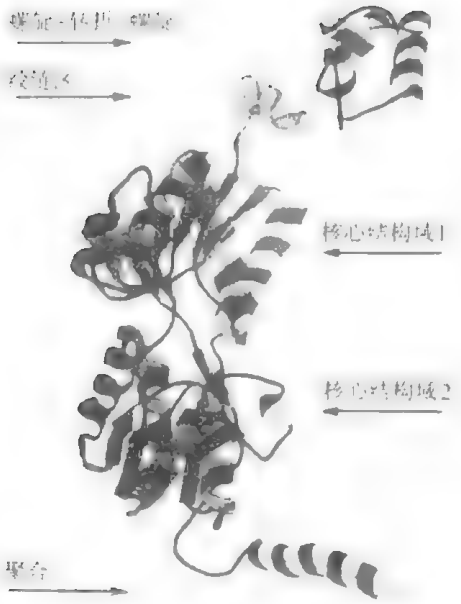
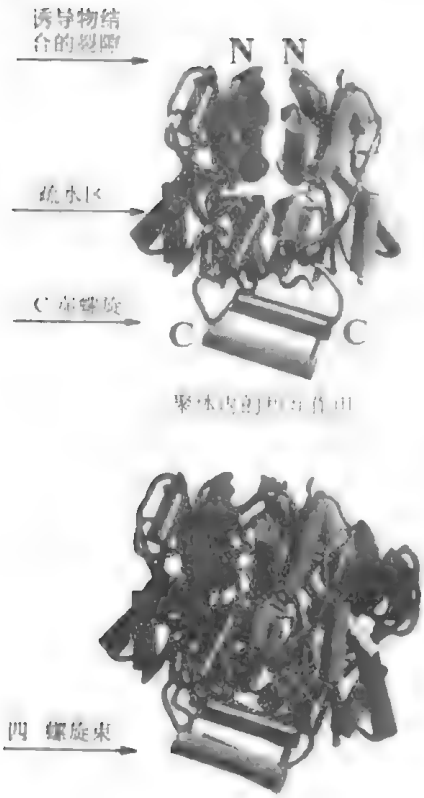


图 7-6 lac 阻抑蛋白亚基的主要结构域



两个二聚体形成一个四聚体

图 7-7 乳糖操纵子核心段结构模型
该结构模型主要说明了四聚体中单亚基间如何相互作用。

图 7-7 示意了乳糖操纵子阻抑蛋白四聚体另一种结构模型，与图 7-6 所示的结构模型不同。图 7-7 所示的结构模型强调了两个二聚体的作用。二聚体的主体在核心段单亚基的 N 端区间结合很松散，二聚体间还形成了诱导物结合裂隙及一疏水区(图 7-7 上图)。每个亚基的 C 段伸出，形成平行的螺旋。头段参与形成上图的 N 端区。两个二聚体一起相互作用形成四聚体(图 7-7 中图)，它们通过 C 端的 4 个螺旋捆扎在一起。突变点用下图结构中的珠子表示。*lacI*^{s} 突变位于两个区域：上部中间的珠子示意影响二聚体相互作用的突变。中部靠近两侧的珠子示意位于诱导物结构裂隙中的突变。*lacI*$^{-}$ 突变影响单亚基间的聚合，该突变也位于两个区域：中下部中间的珠子示意阻止二聚体形成的突变，底部 C 端两个平行螺旋间的珠子示意阻止二聚体形成四聚体的突变。

早期工作建立的阻抑蛋白模型认为，头段相对独立于核心段。与半个结合位点的阻抑蛋白同完整阻抑蛋白以同样方式结合操纵基因 DNA。然而，不完整的阻抑蛋白与 DNA 的亲和力较完整的阻抑蛋白低很多数量级。形成这种差异的原因是二聚体形成完整阻抑

蛋白使两个头段能同时与操纵基因结合，这大大地增加了其与操纵基因的亲和力。

与诱导物结合可直接引起阻抑蛋白构型发生改变。四聚体的阻抑蛋白与两分子诱导物结合足以使阻抑蛋白从操纵基因上释放出来。阻抑蛋白与诱导物结合，使头段与核心段的相当定位改变，结果二聚体的两个头段不再同时与 DNA 结合。这使得多亚基阻抑蛋白失去优势，降低了与操纵基因的亲和力。

4. 操纵基因 *O1*、*O2* 和 *O3* 对阻抑蛋白阻抑能力的影响

阻抑蛋白与诱导物结合引起阻抑蛋白二聚体别构转换。那么为什么形成完整阻抑需要四聚体呢？在四聚体阻抑蛋白中，每个二聚体都有两个头段，两个头段在一起能结合一个操纵基因序列，这使整个阻抑蛋白能同时结合两个操纵基因位点。事实上，在乳糖操纵子起始区中还有另外两个操纵基因位点。基本操纵基因 *O1* 正好位于 *lacZ* 基因的开始区，它与阻抑蛋白亲和力最强。两个弱的操纵基因（有时称为假操纵基因）位于基本操纵基因的两侧，*O2* 位于开始点下游 410 bp 处，*O3* 位于开始点上游 83 bp 处。当 Lac 阻抑蛋白与 *O1* 结合时，会同时与另一个弱操纵基因结合，这使结合点之间的 DNA 形成一个环。

与额外操纵基因的结合可影响阻抑效率。去除下游操纵基因（*O2*）或上游操纵基因（*O3*）中的任意一个，都会使阻抑效率降低 2～4 倍。然而，如果同时删除 *O2* 和 *O3*，阻抑效率降低 100 倍。表明阻抑蛋白结合另两个操纵基因中一个的能力对阻抑作用的形成也像结合 *O1* 一样重要。至于同时结合怎样增加阻抑，为什么会增加阻抑仍不清楚。

5. 阻抑蛋白对 RNA 聚合酶功能的影响

我们了解得最多的是阻抑蛋白与 *O1* 的结合。最初认为阻抑蛋白与操纵基因结合阻碍了 RNA 聚合酶与启动子的结合。后来发现阻抑蛋白和 RNA 聚合酶可同时与 DNA 结合，且阻抑蛋白与 DNA 的结合能增强 RNA 聚合酶结合 DNA 的能力。但是结合着的 RNA 聚合酶不能起始转录。RNA 聚合酶单独结合 *lac* 启动子的平衡常数是 $1.9\times10^{7}\ L\cdot mol^{-1}\cdot S^{-1}$。阻抑蛋白存在时，该常数增加 2 个数量级达 $2.5\times10^{9}\ L\cdot mol^{-1}\cdot S^{-1}$。

这个常数与操纵子的诱导有什么关系呢？这个常数较大时，则意味着被阻抑蛋白结合的启动子与 RNA 聚合酶结合的机会增加 100 倍。RNA 聚合酶与阻抑蛋白同时与 DNA 结合，使得一旦诱导，就能开始转录，无须等待 RNA 聚合酶被“俘获”。

阻抑蛋白实际上使 RNA 聚合酶贮存在启动子上。RNA 聚合酶-阻抑蛋白-DNA 复合物在关闭阶段被阻止发生作用。当加入诱导物后，释放出阻抑蛋白，关闭的复合物转变成开放的复合物，起始转录。阻抑蛋白的综合作用加速了诱导过程。

阻抑蛋白增强 RNA 聚合酶与 DNA 结合的这种模式是否也存在于其他操纵子系统中呢？RNA 聚合酶、阻抑蛋白和启动子/操纵基因区之间的相互作用在不同的操纵子系统中不同，因为有的操纵子中，操纵基因不与启动子位于同一区，互不重叠。例如，在 λ 噬菌体中，操纵基因位于启动子上游区，阻抑蛋白的结合阻碍 RNA 聚合酶的结合。因此，在所有操纵子系统中，结合的阻抑蛋白与 RNA 聚合酶不以同样的方式相互作用。

五、蛋白质与 DNA 的特异结合作用

在体外实验中，可以测定到蛋白质与 DNA 的结合，并可从其亲和力推测调控相互作用的本质。存在的困难是我们不了解体内 DNA 的有效浓度。由于一些因素影响，有效浓度与重量/体积的浓度不同。例如，当多价阳离子中和 DNA 上约 90% 的电荷，使分子拥挤在一起时，以及核酸的结构变为更密实时，有效浓度增加。DNA 有效浓度降低的主要原因是 DNA 结合蛋白难以接近 DNA。

测定有效浓度的一种方法是比较反应在体内和体外速度，这种反应依赖于 DNA 的浓度，这可用两个 DNA 的分子间重组来完成，为了提供对照，接着进行同样的一个分子内重组反应，即在同一 DNA 分子上有两个重组位点。假设分子内、体内和体外反应的浓度相同，这样分子间重组率和分子内重组率的比率的差异是由体内有效浓度的改变引起的。这样比较的结果表明，体内 DNA 有效浓度低 10 倍以上。

但是所有 DNA 结合蛋白没有足够速度找到靶分子，仅靠扩散显然速度不够。也许，所有蛋白质与特异 DNA 序列间亲和力很高，而与随机 DNA 序列间亲和力很低。但大量低亲和力位点则与小量高亲和力位点竞争四聚体阻抑蛋白。在大肠杆菌基因组中，仅有一个高亲和力的位点，这就是操纵基因。除操纵基因外，大肠杆菌基因组中其他 DNA 都是阻抑蛋白的低亲和力位点，因此有 4.2×10^6 个低亲和力位点。大量的低亲和力位点意味着，即使没有特异的结合位点，几乎所有的阻抑蛋白仍能与 DNA 结合，而不游离在溶液中。

阻抑蛋白与 DNA 的结合可用下列公式表示：

$$K_A=\frac{[\text{阻抑蛋白-DNA}]}{[\text{游离阻抑蛋白}][\text{DNA}]}$$

式中，[阻抑蛋白-DNA]是结合 DNA 的阻抑蛋白浓度，[游离阻抑蛋白]是游离阻抑蛋白浓度，[DNA]是非特异性结合位点浓度。

游离的阻抑蛋白在总阻抑蛋白中占多大比例呢？重排上述公式，可以发现游离阻抑蛋白比例为：

$$\frac{[\text{游离阻抑蛋白}]}{[\text{阻抑蛋白-DNA}]}=\frac{1}{K_A\cdot[\text{DNA}]}$$

在乳糖操纵子系统中，其结果是：

1）非特异结合常数 $K_A=2\times10^5$ (L/mol)。

2）在 10^{-15} L 的细菌中，非特异结合位点的浓度是 4×10^6，而 DNA 浓度是 7×10^3 mol/L(浓度很高)。

将这些值代入公式，得出：

$$\text{游离阻抑蛋白}:\text{结合阻抑蛋白}=10^{-4}$$

因此，除了 0.01% 外，所有的阻抑蛋白都与 DNA 随机结合。

因为每个细胞中约有 10 分子阻抑蛋白，这等于说细胞内没有游离的阻抑蛋白。分析阻抑蛋白与操纵基因相互作用时，必须关注 DNA 上的阻抑蛋白比例，操纵基因上的单个高亲和力结合位点与大量低亲和点位点竞争。在这个竞争中，操纵基因及随机 DNA 的

缔合常数的绝对值并不重要，重要的是 K_{sp}（结合特异位点常数）与 K_{nsp}（结合任何随机DNA序列常数）的比例，即特异性的强弱。

我们能通过比较特异结合位点和非特异性结合位点平衡常数来确定影响调节蛋白饱和其靶位点的能力的参数。重要的参数有：

1）基因组的大小，基因组大会使特异位点显得稀少；

2）蛋白质的特异性，蛋白质的特异性强可抵消DNA量大的作用；

3）所需蛋白质的量，随基因组中DNA总量的增加而增加，随特异性的降低而降低。

为了解乳糖操纵子阻抑蛋白与其操纵基因结合的强度和特异性，可对乳糖操纵子阻抑蛋白/操纵基因结合平衡常数与乳糖操纵子阻抑蛋白/一般DNA结合平衡常数进行比较（表7-1）。经过比较可以推断出阻抑蛋白在操纵基因与其他DNA之间怎样分配，当加入诱导物后，阻抑蛋白怎样从操纵基因上解离。

表7-1 乳糖操纵子阻抑蛋白与其操纵基因结合平衡常数

DNA	阻抑蛋白	阻抑蛋白+诱导物
与操纵基因结合的平衡常数(L/mol)	2×10^{13}	2×10^{10}
与其他DNA结合的平衡常数(L/mol)	2×10^{6}	2×10^{6}
特异性	10^{7}	10^{4}

阻抑蛋白结合操纵基因的能力为结合同样长度的其他任何DNA能力的约10^7倍。因此，操纵基因含单一高亲和力位点，该位点与任何（随机）低亲和力位点竞争阻抑蛋白时，其竞争能力也高10^7倍。这保证了阻抑蛋白能维持对操纵子的有效控制。

用特异性的这个参数，能计算阻抑蛋白在随机位点和操纵基因上的分布，并用术语操纵基因占有率（occupancy of the operator）来表示这种分布。如果每个细胞内有10分子乳糖操纵子阻抑蛋白，与操纵基因的特异性是10^7，操纵基因96%的时间都与阻抑蛋白结合。特异性的这种规律解释了乳糖操纵子阻抑蛋白与操纵基因相互作用的两个特点：

1）当诱导物与阻抑蛋白结合时，阻抑蛋白与操纵基因的亲和力下降约10^3倍。而与一般DNA序列的亲和力保持不变。因此，此时特异性仅为10^4，这使得操纵基因没有足够的亲和力与大量（4.2×10^6）的低亲和力位点竞争俘获阻抑蛋白。在这种情况下，仅有3%的操纵基因与阻抑蛋白结合。

2）操纵基因与阻抑蛋白亲和力降低20～30倍的突变就形成组成型突变。在染色体内，突变操纵基因的作用会被大量随机位点的作用所淹没。即使操纵基因与阻抑蛋白的特异性降低仅10%，操纵基因对阻抑蛋白的占有率也将下降至50%。

如此大小的亲和力使得在一个未诱导的细胞内，一个阻抑蛋白四聚体通常与操纵基因结合。所有或几乎所有的剩下的阻抑蛋白在其他区域（随机位点）上与DNA结合。在细胞内几乎没有游离的阻抑蛋白四聚体。加入诱导物就会消除阻抑蛋白特异结合操纵基因的能力。结合在操纵基因上的那些阻抑蛋白被释放下来，结合到随机（低亲和力）位点上。因此，在经诱导的细胞内，阻抑蛋白四聚体都"贮存"在随机DNA位点上，在未经诱导的细胞中，一个阻抑蛋白四聚体结合在操纵基因上，而其他的阻抑蛋白分子结合在非特异性位点上。因此诱导的作用是改变阻抑蛋白在DNA上的分布，而不是形成游离的阻抑蛋白。

除掉诱导物后，阻抑蛋白能很快地恢复特异结合操纵基因的能力。因此阻抑蛋白--

定要从DNA上非特异的"贮存"位点移出。这种快速移出的机制是什么呢？操纵基因快速恢复与阻抑蛋白结合所需的时间与阻抑蛋白在DNA非特异位点上经解离与结合多个循环到达操纵基因所需的时间不一致。这说明阻抑蛋白回到操纵基因位点不是通过随机接触找到操纵基因的，而是从DNA上随机位点直接移到操纵基因上的。这与我们以前碰到的RNA聚合酶找到它的启动子的能力的问题是一样的。因此问题的答案也可能相同。阻抑蛋白的移动能通过在位点间的直接转移完成。阻抑蛋白四聚体有多个结合位点可能有助于这种直接转移。

在体内，乳糖操纵子在诱导和阻抑状态表达的实际差异达10^3倍。换句话说，即使没有诱导物时，也有相当于诱导水平时0.1%的基础表达。如果阻抑蛋白更多，表达水平将下降，如果阻抑蛋白更少，表达水平将上升。因此，如果每个细胞中的阻抑蛋白少于10个，就不可能形成严格的阻抑，反之，如果每个细胞中的阻抑蛋白太多，则操纵子的诱导很困难。

六、色氨酸操纵子控制模式——多基因座同时阻抑

乳糖操纵子阻抑蛋白仅作用于*lacZYA*基因簇的操纵基因。然而，一些阻抑蛋白通过结合多个操纵基因控制散在分布的结构基因。其主要例子是色氨酸阻抑蛋白，它控制三套不直接相连的基因。

1）第一套基因是结构基因簇*trpEDBCA*，此控制点上，控制从分支酸(chorismic acid)合成色氨酸(tryptophan)的酶的合成。

2）第二套基因是*aroH*基因，*aroH*基因编码在芳香族氨基酸生物合成共同通路中催化起始反应的三种酶中的一种。

3）第三套基因是*trpR*调节基因，*trpR*调节基因被其自身产物Trp阻抑。因此阻抑蛋白会降低自身的合成。这种控制是自体控制(autogenous control)的一个实例。这种控制在调节基因的控制中相当普遍，有的是正调控，有的是负调控。

*trp*操纵基因序列长21 bp，三个基因座上均有一个该序列，是*trp*阻抑蛋白作用的位点。图7-8为色氨酸操纵基因保守序列。每个操纵基因含明显的(但不是相同的)对称结构。三个操纵基因保守序列中都有*trp*阻抑蛋白与之结合的部位。正是由于每一个基因座上都有一拷贝能被阻抑蛋白识别的特异DNA结合序列，使得几个部位都可结合同一种阻抑蛋白。

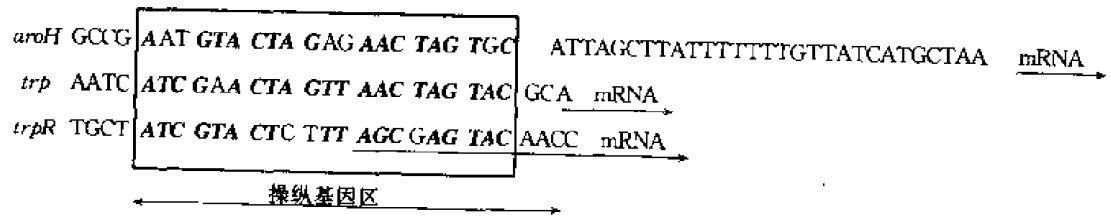


图7-8 三个基因座上的色氨酸阻抑蛋白识别的操纵基因序列

保守碱基用斜体字母表示。

图7-9总结了操纵子当中启动子与操纵基因的相对位置关系。色氨酸阻抑蛋白

(TrpR)识别的散在分布的操纵基因有明显的特点，每一个基因座的操纵基因都在启动子内的不同位置上。色氨酸阻抑蛋白基因(*trpR*)的操纵基因位于－12至＋9之间，而色氨酸操纵子的操纵基因则位于－23至－3之间，*aroH*基因座的操纵基因则在上游更远的区域，位于－49至－29之间。而其他操纵子的操纵基因或者位于启动子的下游(如乳糖操纵子*lac*)，或者位于启动子的上游(如半乳糖操纵子*gal*，其中的天然阻抑作用相当弱)。阻抑蛋白在不同位置的操纵基因上阻抑能力不同，因此确切的阻抑模式可能有差异，但它们的共同点是都抑制RNA聚合酶在启动子上起始转录。

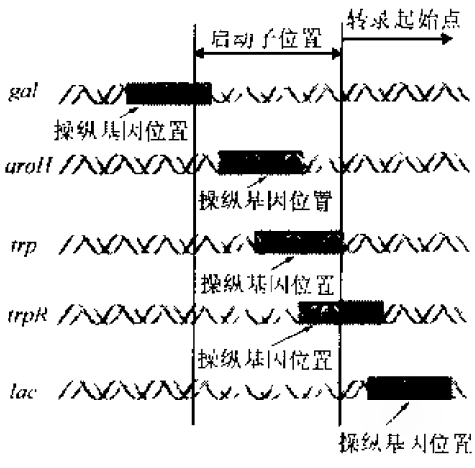


图7·9 各种操纵基因与启动子的相对位置

七、正调控和负调控

正调控和负调控是根据在无调控蛋白的情况下，操纵子应答的结果来确定的。两种控制系统的特性有镜像关系。

在负调控系统中，如果不被阻抑蛋白关闭，基因就能表达。任何干扰基因表达的作用都是负调控，这些调控有一个共同点：即一个阻抑蛋白或与DNA结合，阻止RNA聚合酶启动转录，或与mRNA结合阻止核糖体启动翻译。

负调控是一个万一安全(fail-safe)的机制。即万一调控蛋白失活，调控系统便发挥功能，使细胞能够合成这些酶。因些，很容易理解负调控是如何进化的。最初，系统呈现组成型表达，随后，细胞中产生一种干扰这种表达的机制而给细胞提供了选择优势，这种优势就演化成负调控系统。

受正调控机制调控的基因，仅当活性调控蛋白存在时，才可能表达。正调控中控制特定操纵子的机制正好与负调控对应。正调控与负调控不同，它不妨碍起始，而调控蛋白对正调控来说，则必不可少。调节蛋白与DNA及RNA聚合酶相互作用，协助转录起始。正调节蛋白应答的小分子通常称为激活物(activator)。另一种正调控则进行全局性替换σ因子，改变启动子的选择性，或进行全局性替换抗终止因子，从而改变终止子的识别。

了解正调控的进化过程较了解负调控的更困难，因为甚至在任何调控存在之前，细胞也肯定具有表达这些被调控的基因的能力，可以推测，此控制系统的一些成分的作用肯定发生了变化。也许，原来它是基因表达装置的调节部件，后来变成仅局限于对某些系统或某个特定系统发挥作用。

根据应答调节其表达的小分子的性质，操纵子分为诱导型(inducible)和阻抑型(repressible)两类。只有在小分子诱导物存在时，诱导型操纵子才能发挥作用；只有在小分子辅阻抑物(corepressor)不存在时，阻抑型操纵子才能发挥作用。

描述阻抑型系统中操纵子活性状态的术语是去阻抑(derepressed)，去阻抑与诱导有同样的含义。有时将不能去阻抑的突变操纵子称为超去阻抑(super-derepressed)，与不可诱导型完全对应。

通过利用调控蛋白与小分子诱导剂或辅阻抑物适当的相互作用，无论是正调控还是

负调控都能实现诱导或阻抑。图 7－10 总结了 4 种简单控制的流程。当诱导物使阻抑蛋白失活或使激活物蛋白激活时，实现诱导。当辅阻抑物使阻抑蛋白激活或使激活物蛋白失活时，则实现阻抑。

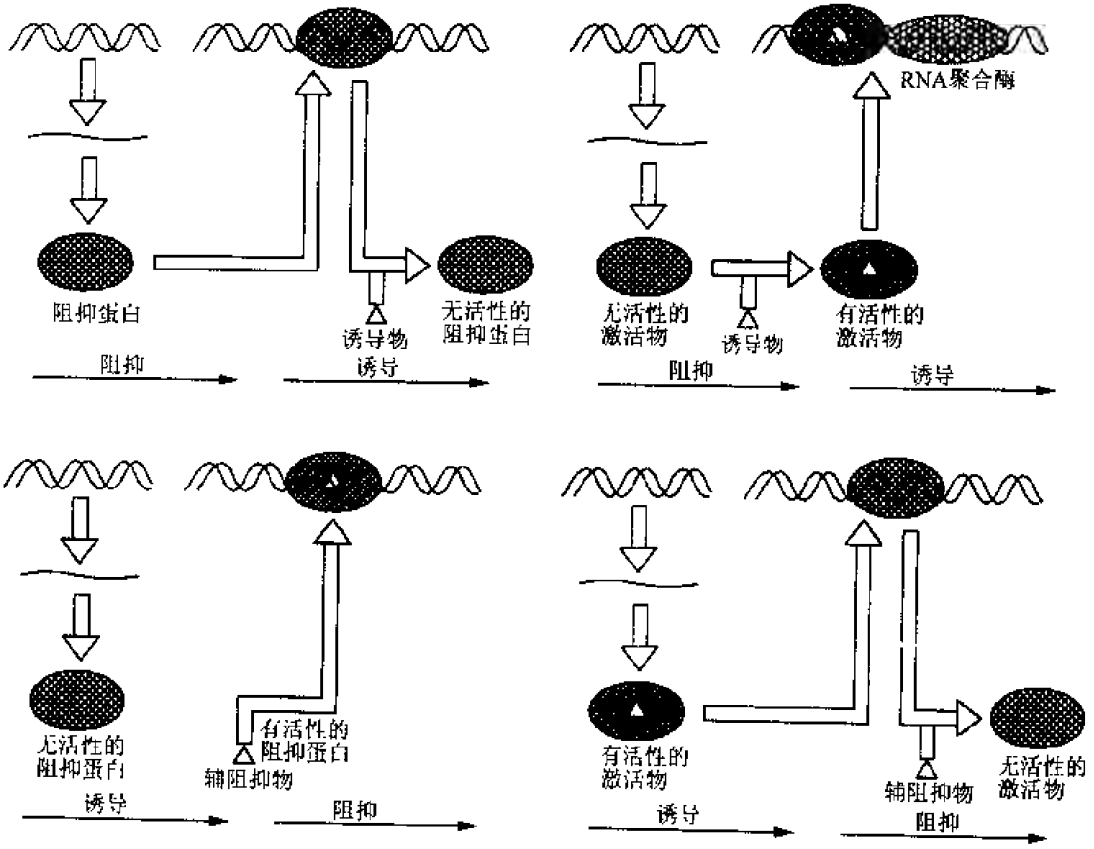


图 7－10　正、负调控控制诱导和阻抑的控制流程

左边两图为负调控，右边两图为正调控。左上和右上两个上图为诱导，左下和右下两个下图为阻抑。

调节蛋白失活的遗传结果能用于区别正调控系统和负调控系统。阻抑蛋白的失活产生代表负调控系统的隐性组成型或去阻抑型表型。另一方面，使激活物失活的突变，形成代表正调控系统的隐性不可诱导型或超阻抑型表型。

色氨酸操纵子是阻抑系统的一个实例。色氨酸是一系列生物合成酶催化反应的终产物。细胞内色氨酸的水平控制色氨酸酶的合成和活性。在这些调控中色氨酸起激活阻抑蛋白的辅阻抑物的作用，这是经典的阻抑机制，实例如图 7－10 中的左下图所示。当环境中的色氨酸丰富时，由于阻抑蛋白与辅阻抑物复合物与操纵基因结合，操纵子被阻抑。当环境中提供的色氨酸不足时，辅阻抑物失活，与操纵基因结合的特异性下降，辅阻抑物贮存 DNA 的其他位点上，操纵子的阻抑被解除。

清除阻抑蛋白能使 *trp* 启动子上转录起始的频率增加约 70 倍。即使在阻抑条件下，结构基因也以低的基础水平或称阻抑水平继续表达。在操纵基因上的阻抑效率较乳糖操纵子低得多。乳糖操纵子基础水平表达仅为诱导水平表达的千分之一。

小分子引起调节蛋白变构也可实现诱导和阻抑。控制调节蛋白的活性也能达到调控

的目的。其中的例子是 OxyR，它是由过氧化氢诱导的一种基因转录激活剂。OxyR 蛋白直接被氧化所激活。因此它提供了一种敏感的氧化应激检测手段。另一个常见的调控信号是调节蛋白的磷酸化。

八、分解代谢产物阻抑作用对启动子的正调控

到目前为止，我们一直把启动子看作一段 DNA 序列，它能与 RNA 聚合酶结合，从而起始转录。但是有一些启动子，如果没有辅助蛋白的协助，RNA 聚合酶并不能在其上起始转录。因为要打开转录单位的开关，必须有这些蛋白存在，所以这些蛋白是正调控物。已知几种正调控物，一些被噬菌体编码，一些存在于宿主细胞内。它们发挥作用的一种模型是激活物克服了低效率启动子效率低的问题，这些低效率启动子位于没有很好的保守序列的 -35 区或 -10 区。

1. 分解代谢产物的阻抑作用

在大肠杆菌中，一种蛋白质控制着应答碳源条件的一大套操纵子的活性，这种蛋白质便是发挥作用最广泛的激活物之一。当葡萄糖作为能源时，它能比其他糖优先被利用。因此当大肠杆菌在培养基中发现葡萄糖和乳糖时，它分解利用葡萄糖，而阻抑乳糖的利用。这种选择是通过阻止包括乳糖操纵子、半乳糖操纵子和阿拉伯糖操纵子等的表达来实现的。这种作用被称为分解代谢产物阻抑作用(catabolite repression)。

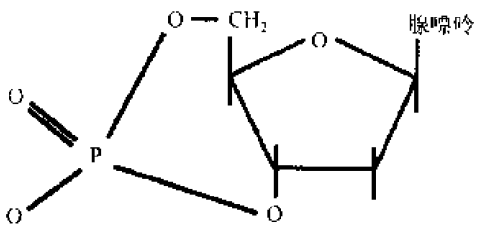


图 7-11　cAMP 结构

分解代谢产物的阻抑作用属于葡萄糖降低细胞内 cAMP 水平的调控系列环节。cAMP 由腺苷酸环化酶(adenylate cyclase)催化合成。腺苷酸环化酶以 ATP 为底物，将核糖环上 3′、5′两个碳原子通过磷酸二酯键连接，形成图 7-11 所示的 cAMP 结构。编码腺苷酸环化酶基因的突变(cya^-)不能应答葡萄糖水平的变化。

2. 分解代谢物激活蛋白(CAP)的功能

cAMP 通过与 *cap* 基因产物结合，使分解代谢物调控的操纵子的表达水平与 cAMP 水平成反比。*cap* 基因产物称为分解代谢物激活蛋白(catabolite activator protein, CAP)。*cap* 基因突变可阻止操纵子的激活，而在葡萄糖缺乏时，这些操纵子能被正常表达。CAP 是一正调控因子，在依赖 CAP 的启动子上起始转录必须有 CAP 参与。只有 cAMP 存在时，CAP 才有活性，cAMP 起到了传统的小分子诱导物的作用。

cAMP 下降，CAP 就不能与控制区结合，RNA 聚合酶就不能启动转录(图 7-12)。因此，葡萄糖降低 cAMP 水平的作用，实际上是使相关操纵子表达所需的调控因子失活，使操纵子不能被表达。

CAP 因子可与 DNA 结合，从每个正常发挥作用的启动子上都能分离到 cAMP·CAP·DNA 复合物。CAP 因子是由二个相同亚基构成的二聚体，每个亚基都含螺旋-转折-螺旋基序，分子量为 22 500 Da，能被单个分子 cAMP 激活。一个 CAP 亚基含有一个 DNA 结

合区和一个转录激活区。

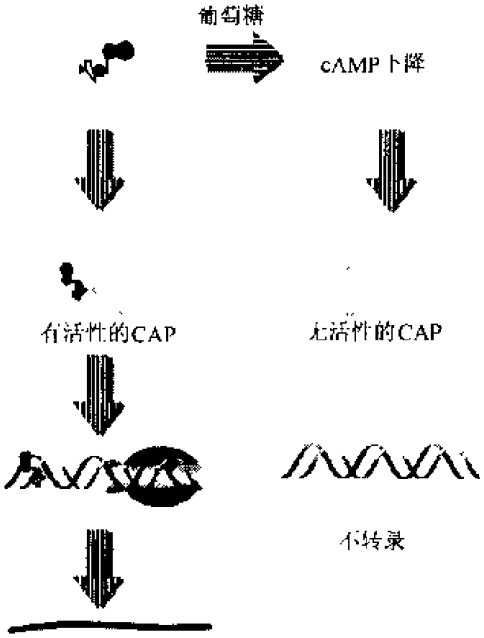


图 7-12 分解代谢物阻抑作用
葡萄糖降低 cAMP 水平，引起分解代谢物阻抑作用。

3.CAP 结合位点的结构

CAP 二聚体与其效应的启动子结合位点长约 22 bp。结合位点保守序列见图 7-13。使 CAP 失去作用的突变通常位于保守的五碱基单位 $\frac{TGTGA}{ACACT}$ 内，这些碱基是识别的基本元件。因为 CAP 二聚体的两个亚基有效地与 DNA 结合，所以与 CAP 结合最强的位点含两个反向重复的五碱基单位。然而很多结合位点缺乏第二个五碱基单位，因此，在这些结合位点上第二个亚基必须与不同的序列结合（如果与 DNA 结合的话）。与 CAP 结合亲和力的差异阐明了为什么在体内不同的基因受 cAMP 激活的水平不同。

4.CAP 结合位点与转录起始点的位置关系

不同操纵子的 CAP 结合位点与转录起始点的相对位置不同。甚至 TGTGA 五碱基单位可处在启动子的不同方向上，有的在上游，有的在中（图 7 14）。图 7-14 的三个例子涵盖了 CAP 结合位点的三种情况：

1）有时 CAP 结合位点位于启动子内。例如 *gal* 基因座，CAP 位点位于 -50 至 -23 bp 处，中心在 -41 bp 处。由于 CAP 结合位点正好位于通常被 RNA 聚合酶保护的区域内，所以很可能仅有一个 CAP 结合，且与 RNA 聚合酶有很紧密的接触。

图 7-13 CAP 结合位点保守序列
左框内为高保守的五碱基单位 TGTGA，右框内为其倒转序列——保守性稍低的 5 碱基单位 TCANA。

2）CAP 结合位点与启动子相邻，例如乳糖操纵子。在乳糖操纵子中，受 CAP 保护的 DNA 区域位于 -72 至 -52 bp 处，中心在 -61 bp 处。很可能有两个 CAP 与其结合。其结合模式为：CAP 主要结合在 DNA 的一个面上，RNA 聚合酶也正好结合在此面上，CAP 与 RNA 聚合酶正好相互挨着。

3）其他操纵子 CAP 结合位点位于启动子上游，如阿拉伯糖操纵子（*ara*）其 CAP 结合位点与一个 CAP 结合，距转录起始点远，位于 -107 至 -78 bp 之间。因 CAP 和 RNA 聚合酶间有另一调控蛋白结合，故 CAP 不能与 RNA 聚合酶接触。

对 CAP 的依赖与启动子本身的效率有关，非 CAP 依赖启动子有好的 -35 序列，但一些启动子没有好的 -10 序列。实际上，我们可以推论，如果启动子含有效的能独立与 RNA 聚合酶作用的 -35 区和 -10 区，CAP 的有效调控将十分困难。

5.CAP 对 RNA 聚合酶的作用

理论上 CAP 激活转录有两种方式，第一种是 CAP 直接作用于 RNA 聚合酶，另一种作用于 DNA，改变其结构，以协助 RNA 聚合酶的结合。实际上，CAP 对 RNA 聚合酶和 DNA 都有效。

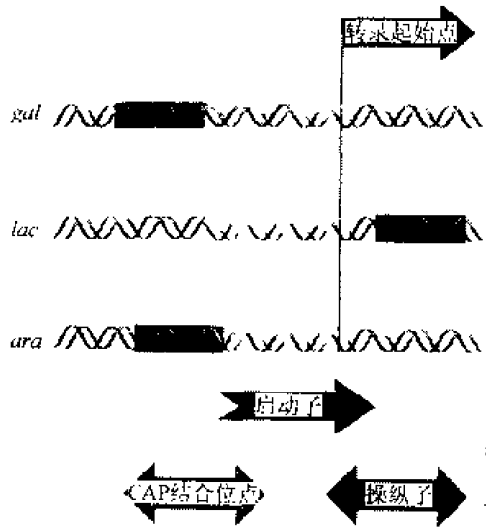


图 7-14 CAP 位点与启动子等的位置关系

大多数启动子上的 CAP 结合位点与 *gal*（中心在 −41 bp）或 *lac*（中心在 −61 bp）相似。如其位置偏离这两处，CAP 激活转录的能力下降。其位置的规律是仅当 CAP 与转录起始点相距数个整双螺旋时，才能激活转录。表明 CAP 必须与 RNA 聚合酶在同一个面上才有活性。

当 RNA 聚合酶的 α 亚基 C 末端存在缺失时，转录似乎仍能正常进行，但失去受 CAP 激活的能力，转录则不能正常进行。这表明 CAP 直接作用于 RNA 聚合酶的 α 亚基。使用只有一个亚基且具有正常转录—激活区功能的 CAP 二聚体进行的实验表明，当与 *lac* 启动子结合时，只需要与起始点较近的亚基的激活区，这可能是它能与 RNA 聚合酶接触，这也解释了为什么 CAP 结合位点在转录起始点上、下游都可发挥作用，没有方向性的要求。CAP 具有两个亚基也保证了不管在上游还是下游，激活区都能与 RNA 聚合酶接触。CAP 激活区含有一个暴露的小环，其中的点突变可阻断 CAP 与 RNA 聚合酶相互作用。

RNA 聚合酶与 DNA 的结合受 RNA 聚合酶与 CAP 的相对位置影响。当 CAP 结合在启动子内（如在 *gal* 中）时，能加快从关闭型复合物到开放型复合物的转换速度。当 CAP 结合位点与启动子相邻（如在 *lac* 中）时，能加快起始结合形成关闭型复合物的速度。这表明相互作用的确切效果与蛋白质在各启动子处的几何形状有关。在 *ara* 启动子处的相互作用可能不同，有更多蛋白质参与。

6.CAP 的结合对 DNA 构型的影响

CAP-DNA 复合物的结构很有意思，DNA 呈弯曲状态。蛋白质与 DNA 结合时，将 DNA 双螺旋变弯。

可以对弯曲的内容（碱基成分）和位置进行测定。制备一个靶序列二聚体，用不同的限制酶进行酶切，得到一系列片段，每个片段含有一个单体长度的 DNA，那么蛋白结合位点位于每个片段的不同位置上。

弯曲的位置不同，片段在凝胶电泳中移动速度不同，如果没有弯曲，所有片段移动速度相同。当弯曲位于 DNA 片段中心时，对运动阻碍最大，泳动速度最慢；弯曲位于一端时，对运动阻碍最小，泳动速度最快。

将电泳迁移率对酶切位点作图，最大迁移率在曲线是最低点，在这个位置上，酶切序列与弯曲点紧挨着。

CAP 与 *lac* 启动子相互作用，弯曲点位于二重对称的中心。弯曲相当严重，弯曲度大于 90°（图 7-15）。所以 CAP 结合后，DNA 双螺旋结构有很大变化。弯曲的机制是：TGTGA 保守序列内产生一个尖纽结，每个拷贝中的两个纽结呈回文结构，引起超过 90°的弯曲。弯曲对转录可能有一些直接作用，但是更可能的情况是弯曲使 CAP 能与启动子上的 RNA 接触。

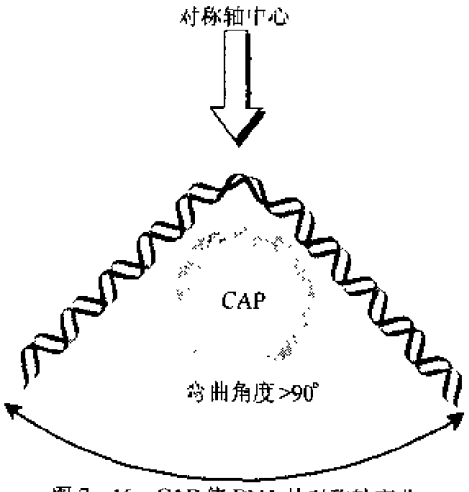


图 7-15　CAP 使 DNA 从对称轴弯曲

在多种启动子中 CAP 激活转录的精确过程可能不同，但是它们达到相同的总目标：由于给细胞提供了足够的葡萄糖，关闭那些不必要的代谢通路。此外，这点也是负调控和正调控的协调控制，涵盖了调节蛋白对散在分布的基因座多个结合位点的结合。

第三节　转录后加工的调控[1,7]

原核生物转录未完成就开始进行翻译，因此，实际上调控也在同时进行，不能分开。但我们仍将 mRNA 切割与 rRNA 切割的调控放在此节叙述。

一、经 mRNA 切割进行的调控

细菌 mRNA 几乎都以转录出的最初形式进行翻译，而无须加工。细菌的 mRNA 大多数是多顺反子，也有一些是单顺反子。多顺反子无须切割成单个基因的 mRNA，就可以进行翻译。但也有例外，如 T7 噬菌体，早期和晚期基因转录成多顺反子 mRNA，然后 mRNA 分子在几个位点切割释放出各个基因的 mRNA。这种表达形式也存在于 T3 噬菌体中。T7 噬菌体的早期区域的转录见图 7-16。在启动子 *A1*、*A2* 和 *A3* 中的某一个上启动后，转录终止于约 7000 bp 后的终止子 *t* 处。启动子与终止子间有 6 个基因，RNA 酶 III 在转录产物内各顺反子交界处切割，释放出 5 个 mRNA 分子，其中 4 个为单顺反子，1 个为双顺反子。对切割形成的 mRNA 分子 5′和 3′末端序列进行测定，并与已知的 T7 噬菌体的 DNA 序列进行比较，确定了 RNA 酶 III 切割位点的性质。切割位点的共同特点是都拥有双链结构。T7 上所有的 RNA 酶 III 切割位点形状都是含一未配对环的双螺旋发夹结构。如图 7-17 所示，所有位点都被 RNA 酶 III 切割，形成的发夹结构是酶的识别所需要的唯一特征。

T7 噬菌体为了翻译双顺反子 *1.1*/*1.2*mRNA，必须在 *1.2* 基因的末端进行切割。当切割被阻止时，*1.1* 和 *1.2* 编码区都不能翻译。其原因是较长的 mRNA 形成二级结构阻止了 *1.1* 编码区翻译的起始，*1.1* 区翻译失败则阻止了 *1.2* 区的翻译。

RNA 酶 III 影响翻译的另一个例子是 RN 酶 III 对含基因 *int* 的 λ 噬菌体 mRNA 翻译的控制。λ 噬菌体的这个 mRNA 通常不翻译，蛋白质 Int 由另一 mRNA 指导合成。在位于 *int* 编码区下游，称为 *sib* 的位点突变，使蛋白质 Int 从通常无活性的 mRNA 上翻

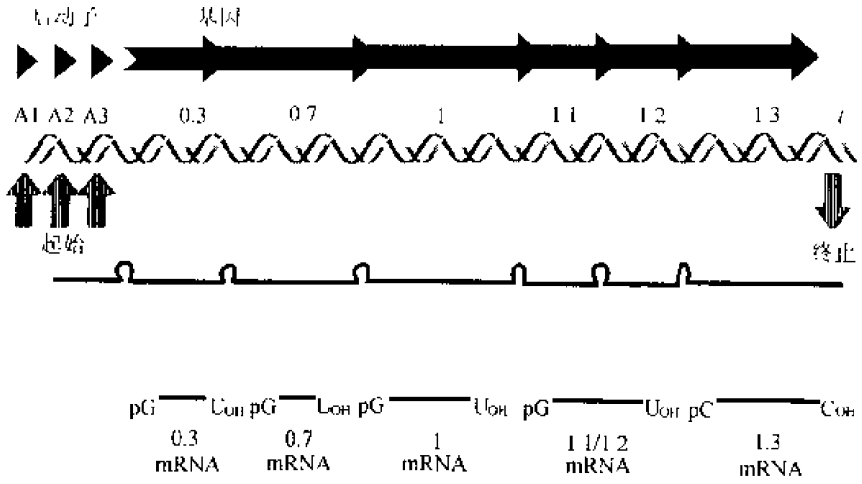


图 7-16 T7 噬菌体早期区转录切割示意图

先转录成一条 RNA，然后被 RNA 酶 III 切割成 5 条 mRNA。

译。*sib* 突变阻止 RNA 酶 III 识别靶位点。而野生型 mRNA 上此位点的切割也许可使 3′末端暴露，使其被另一核糖核酸酶降解。mRNA 的这种不稳定性阻止了蛋白质 Int 的合成。*sib* 及 *rnc* 突变都有相同的效果，都阻止 RNA 酶 III 的切割，因此稳定了 mRNA，使其被翻译。

在 T7 噬菌体的 *1.1/1.2* 基因和 λ 噬菌体 *int* 系统中，mRNA 的翻译能力与 RNA 酶 III 在特定位点上的切割之间有一定的关系。切割事件对翻译功能的影响，主要是改变一些具有重要功能的 RNA 结构，影响的方面主要有 RNA 的翻译起始区，使 RNA 易于降解，或使 RNA 能被剪接(真核生物中)等。

二、通过切割释放原核和真核 rRNA

由于真核生物的 rRNA 初级转录物经过一系列中间过程转变为最后的 rRNA 分子的成熟过程速度相对较慢，所以对其成熟过程了解地较多。主要的加工中间产物能从各种细胞分离到，其中对哺乳动物的加工途径特点了解地最清楚。

哺乳动物 rRNA 的初级转录产物是 45 S RNA，其中含 18 S 和 28 S rRNA 的序列，但是长度几乎是两者之和的两倍。它含有在转录过程中及转录后加上的大约 110 个甲基。甲基在前体加工到成熟 rRNA 过程中被保留下来。在成熟的 18 S rRNA 中有大约 39 个原有的甲基，另外 4 个是后来在细胞浆中加上去的。在 28 S rRNA 上有约 74 个甲基，都是原有的。表明甲基化是用于确定初级转录是否成熟进入 rRNA 的区域。

因为 RNA 分子中间产物有不同的大小，从中已经发现了不止一种成熟途径。假设使用小量次序不同的切割位点，所有的途径都是一致的。在人 HeLa 细胞和鼠 L 细胞中发现了两种哺乳动物加工途径。在每种途径中，在 18 S 基因和 5.8 S 基因间的间隔内及 5.8 S 和 28 S 序列的间隔内，18 S 基因的 5′侧都有切割位点。两种途径中所利用的切割位点的次序不同。还可能存在其他的差异，一种类型的细胞中也曾发现有多种途径。

细菌也需要切割 rRNA 前体形成 rRNA。大肠杆菌有 7 个操纵子编码 rRNA，这 7 个

操纵子被命名为 *rrnA*～*G*，它们在染色体中不紧紧连在一块。rRNA 序列是保守的，但是保守性怎样维持仍不清楚。所有 *rrn* 操纵子的组织结构相同，都含有所有三种 rRNA 分子，顺序为 16 S－23 S－5 S(图 7－18)。在转录单位中两个大的 rRNA 的排列次序与真核生物转录单位中的次序相同，而在同一转录单位内有 5 S RNA 参与，是原核生物所特有的。这些操纵子的每一个都被转录成一个 RNA 前体，然后切割形成成熟的 rRNA 分子。当切割反应被阻断时，前体 30 S RNA 富集。

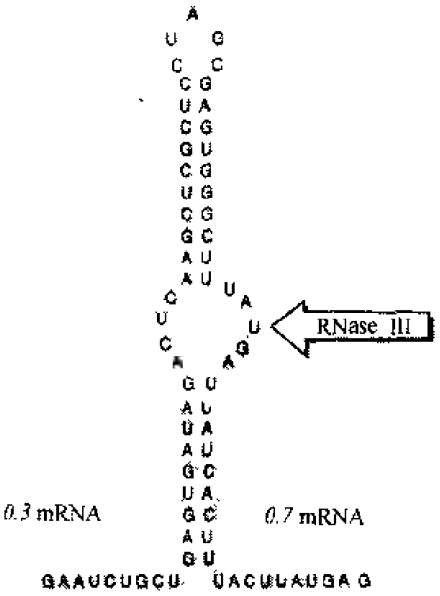


图 7－17 RNA 酶 III 对 T7 RNA 的切割

本图示意了 RNA 酶 III 对 T7RNA 的切割。切割位点位于含未配对泡的短的发夹结构。

所有 *rrn* 操纵子都有双启动子结构。第一个启动子 *P1* 位于 16 S rRNA 序列起始点上游的约 300 bp 处。也许它是主要的启动子。各 *rrn* 操纵子转录单位的头 150 bp 序列各不相同。在这段区域内，距 *P1* 约 110 bp 处是第二个启动子 *P2*。16 S 和 23 S rRNA 两侧的间隔序列是保守的。

在 16 S rRNA 和 23 S rRNA 的序列之间是一个 400～500 bp 的转录间隔区，这个间隔区内有一段或几段编码 tRNA 的序列。如图 7－18 所示，有 4 种 *rrn* 操纵子间隔区的 tRNA 为 $tRNA_2^{Glu}$。另 3 种 *rrn* 操纵子此间隔区含两个 tRNA 分子($tRNA_1^{Ile}$ 和 $tRNA_{1B}^{Ala}$)。因此，根据间隔区 tRNA 序列，可以区别两类 rRNA 操纵子。tRNA 从前体 RNA 中切下后，剩下的间隔区被降解。

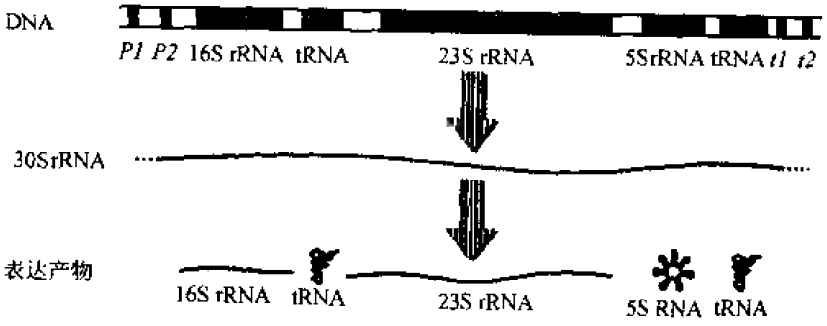


图 7－18 *rrn* 操纵子结构

rrn 操纵子含有 rRNA 和 tRNA 基因，每一功能序列与下一功能序列被转录间隔区分开，前导区和尾区的长度由所用的启动子和终止子决定。每个 RNA 产物都必须在两侧切割后从转录物中释放出来。

在 7 种 *rrn* 操纵子中，还有一个可变区。在其中的一些操纵子中，前体的最后部分编码 5 S rRNA。而在另一些操纵子中，还有额外的 tRNA 序列。例如，在 *rrnC* 操纵子中，3′端有两个基因 $tRNA^{Trp}$和 $tRNA_1^{Asp}$。$tRNA^{Trp}$基因的存在，使得大肠杆菌对这个特殊的 *rrn* 操纵子必不可少了，因为它是这个基因的唯一拷贝，这样，在蛋白质合成中提供了利用色氨酸的独特能力。

负责 rRNA 加工的酶是 RNA 酶 III。在 *rnc*⁻ 细胞中，不出现成熟的 rRNA，而含

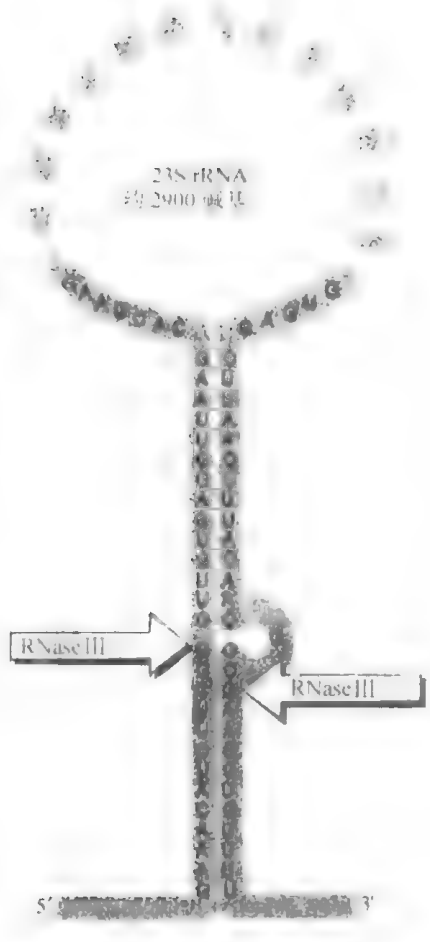


图 7-19 RNA 酶 III 对 rRNA 的切割
rRNA 的 5′和 3′端区形成双螺旋茎，在此茎两侧，RNA 酶 III 进行切割，图中仅示意了 23 S rRNA 上的切割位点，对 16 S rRNA 的切割方式与此相同。

16 S、23 S 和 5 S 序列的 30 S 的前体聚集。这种前体能在体外被 RNA 酶 III 切割形成单个成熟的 RNA 分子前体。

16 S 和 23 S rRNA 的前体称为 p16 和 p23。由于每个前体在 5′端和 3′端留有多余序列，因此都比核糖体中成熟的 rRNA 分子略长。

形成 p16 和 p23 RNA 末端的机制相同。两者的初级转录物中末端都互补，碱基配对形成含产物 5′和 3′末端区的双链结构，这种作用的特点是互补序列间的距离很大，p16 为约 1600 个核苷酸，p23 为约 2900 个核苷酸。如图 7-19 所示，RNA 酶 III 在 p16 或 p23 分子双链区两侧切割，同时产生 p16 和 p23 分子的 5′端和 3′端。

p16 和 p23 分子上 RNA 酶 III 切点间及这些切点与 T7 位点均不存在同源序列。因此 RNA 酶 III 特异性地切割 RNA 时，除了已知切割位点附近需要双链结构外，其它特点仍不清楚。由于切割由 RNA 酶 III 完成，切割后并不释放 16 S rRNA 或 23 S rRNA 实际的 5′末端和 3′末端，因此仍需进一步的加工。对进一步加工过程的了解仍不多，*rrn* 转录单位中的 tRNA 序列的加工是由 RNA 酶 P 及其他 tRNA 加工酶完成的。

第四节　翻译水平的调控[1,8,9]

翻译水平的调控包括对构成蛋白质合成装置的蛋白质的编码基因的操纵子进行控制。采用操纵子可对一组结构基因进行协调控制，实际上这仍然是通过转录水平的调控实现对翻译水平的影响。但是，还有其他的调控方式。例如，可在翻译水平上进行调控，使得操纵子中的各个基因的表达量存在差异。

一、翻译过程中的自体调控

1. 阻抑蛋白对翻译起始的调控

在几个不同的控制系统中，翻译水平控制的机制相似。即通过阻抑蛋白与 mRNA 的特定区域结合，抑制核糖体对翻译起始区的识别。这与阻抑蛋白与 DNA 结合阻止 RNA 聚合酶利用启动子开始转录的调控机制相似。这种调控最常见的作用形式是，调控蛋白与含有起始密码子 AUG 序列直接结合，从而阻止核糖体结合。

表 7-2 列举了一些翻译阻抑蛋白及其作用对象(靶基因)。一个传统的例子是 RNA 噬菌体 R17 的外被蛋白(coat protein)，它与噬菌体 mRNA 核糖体结合位点内的发夹结构

结合。另一个相似的例子是 T4 噬菌体的 RegA 蛋白，它与几种 T4 噬菌体早期 mRNA 上的含有 AUG 起始密码子的一共有序列结合。

表 7－2　与 mRNA 翻译起始区内序列结合抑制翻译的阻抑蛋白

阻抑蛋白	靶基因	作用位点
R17 外被蛋白	R17 复制酶	核糖结合位点的发夹结合
T4 RegA	早期 T4 mRNA	包括起始密码子的各种序列
T4 DNA 聚合酶	T4 DNA 聚合酶	SD 序列
T4 p32	基因 32	单链 5′端前导序列

2. mRNA 二级结构对翻译的调控

另一种形式的翻译调控是：转录一个顺反子需要二级结构的改变，而这种二级结构的改变又依赖于前一个顺反子的翻译。RNA 噬菌体的蛋白质合成过程中，存在这种调控作用。RNA 噬菌体的顺反子总是按一定的顺序表达。如图 7－20 所示，噬菌体 RNA 形成一种二级结构，在此二级结构中，仅一个暴露的翻译起始序列可以利用，第二个翻译起始序列由于与 RNA 上的其它区的碱基配对而不能被核糖体识别。然而，第一个顺反子的翻译破坏原有的二级结构，使核糖体能与下一个顺反子的翻译起始区结合。因此，在这个 mRNA 中，二级结构对翻译进行调控。

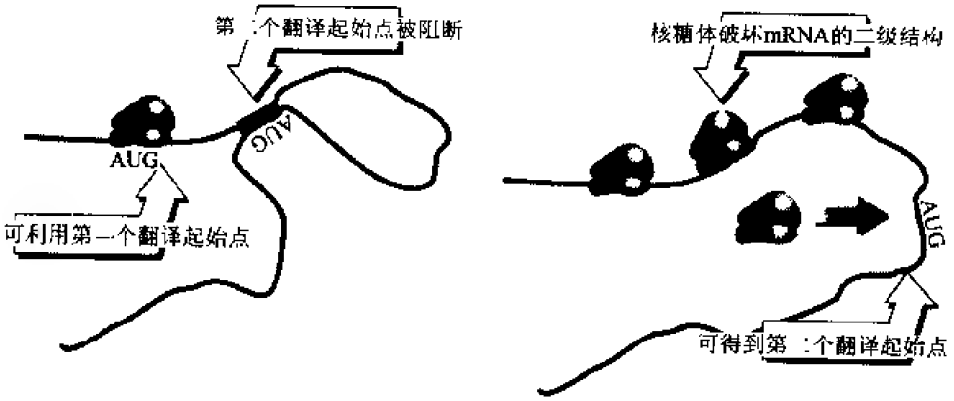


图 7－20　mRNA 二级结构对翻译起始的调控

RNA 噬菌体中，由于 RNA 形成二级结构，仅有一个翻译起始位点可以利用，但翻译第一个顺反子会改变 RNA 的构型，使得第二个翻译起始位点也暴露出来，可以被利用。左图中仅一个翻译起始位点可以利用，右图中第一个顺反子翻译使第二个翻译起始位点暴露出来。

3. 参与构成基因表达装置的蛋白质的合成的自体调控

约有 70 种蛋白质参与构成细菌基因表达的装置。核糖体蛋白是其中最主要的成分，它和其他辅助蛋白一起参与蛋白质合成。RNA 聚合酶亚基及其辅助因子也是其中的成分。编码核糖体蛋白质、蛋白质合成因子和 RNA 聚合酶亚基等蛋白质的基因混合组成几个操纵子。在大肠杆菌中，这些蛋白质大多数仅有一个基因编码。

协调控制保证了这些蛋白质合成适当的量以满足细菌生长的需要。当细菌生长速度很快时，细菌将尽更大的努力来制造组成基因表达装置的成分。有一系机制用来调控编码这种装置的基因表达，从而保证蛋白质的合成水平与 rRNA 的量相当。

图 7-21 总结列出了 6 种操纵子的组织方式。编码核糖体蛋白质的基因有一半分布在紧靠在一起的 4 个操纵子中，它们是 *str*、*spc*、*S10* 和 *α*。其余的两个操纵子 *rif* 和 *L11* 一起位于其他位置上。

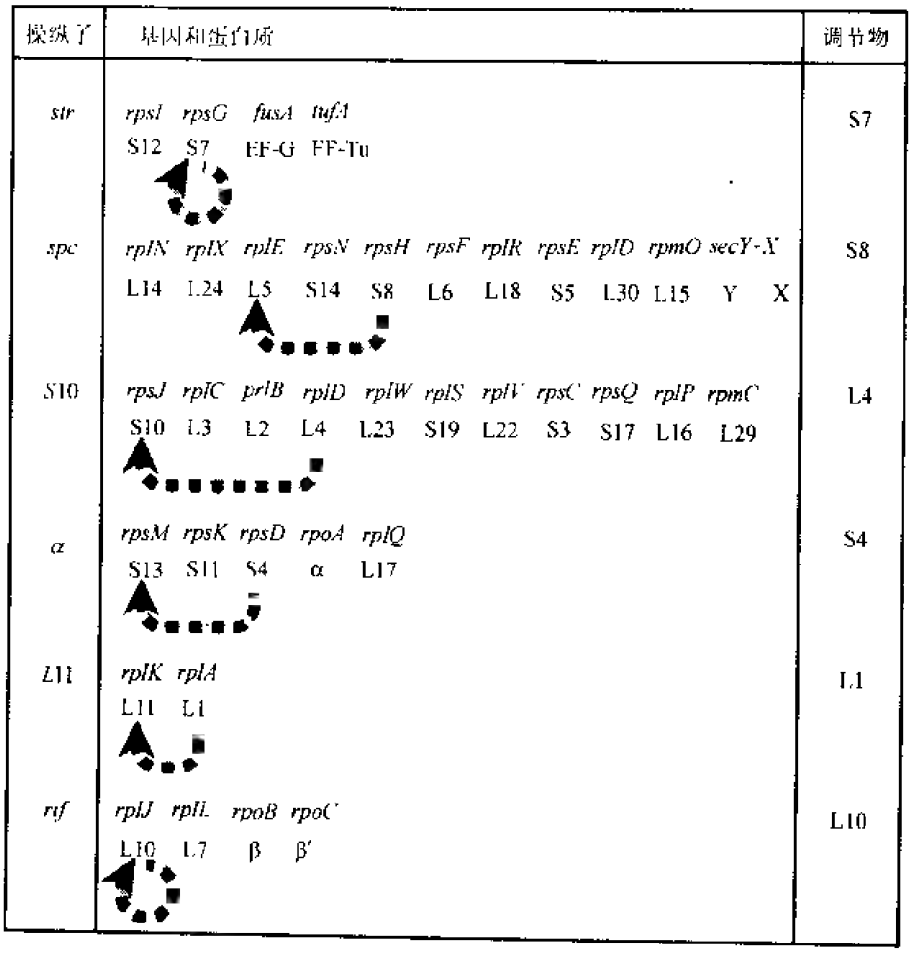


图 7-21 基因产物与蛋白质合成装置有关的操纵子

图中示意编码核糖体蛋白、蛋白质合成因子及 RNA 聚合酶亚基的操纵子及其自体调控，参与调控的蛋白用阴影表示。

每个操纵子都含数个基因。*str* 操纵子具有编码核糖体小亚基的一些蛋白质及 EF-Tu 和 EF-G 的基因。*spc* 和 *S10* 操纵子具有编码组成核糖体大亚基和小亚基的蛋白质的基因。*α* 操纵子具有编码组成核糖体大亚基和小亚基的蛋白质的基因，同时还具有编码 RNA 聚合酶 α 亚基的基因。*rif* 操纵子具有编码核糖体大亚基上的蛋白质及 RNA 聚合酶的 β 和 β′亚基的基因。

绝大多数核糖体蛋白(除了 L7/L12 在每个核糖体中有 4 个分子外)在每个核糖体中只有一个分子，因此它们的表达必须与 rRNA 相协调。对于延伸因子 EF-Tu，每个细胞中

Tu 的分子数大约是核糖体数的 10 倍。RNA 聚合酶的每种亚基数要比核糖体数目少一些。由于这些基因混合编组在不同的操纵子中，因此必然存在一种协调机制，因上述需要不同表达量的基因各得其所。

图 7－22 中所有操纵子的共同点是操纵子的表达都受操纵子自身的一些基因产物的调控。一些基因编码的调控蛋白的调控对象之一是基因本身。一种蛋白质或 RNA 调控其自身的表达，这就是自体调控。r－蛋白操纵子中，调控蛋白抑制操纵子中一组相邻基因的表达，这便是自体负调控的例子。

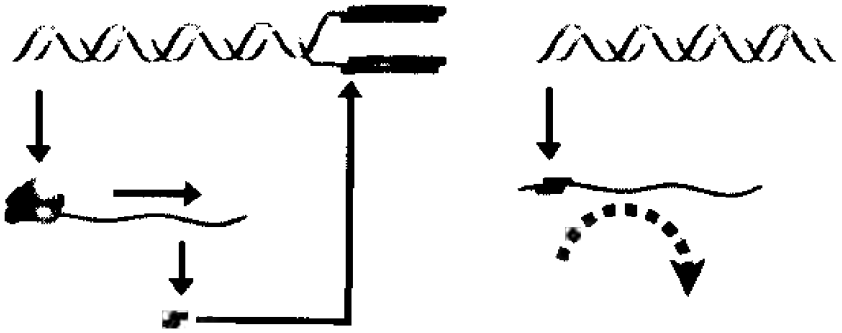


图 7－22　过量的基因 *32* 的蛋白 p32 与自身的 mRNA 结合抑制核糖体翻译起始
左图示意 p32 更易与单链 DNA 结合，当仍存在有其结合位点时，其合成仍继续进行着。右图说明当没有单链 DNA 时，p32 与其自身 mRNA 的起始密码子周围富含 A-T 的区域结合，阻止翻译起始。

蛋白质的富集会抑制其自身及一些相关基因产物的进一步合成。这种作用常见于多顺反子 mRNA 翻译水平的控制。每种调节物都是直接与 rRNA 结合的核糖体蛋白。在翻译中的作用取决于结合自身 mRNA 的能力。mRNA 上这些蛋白的结合位点位于翻译起始序列内或其附近，或通过构型改变从而可能影响翻译起始区发挥作用，例如，*S10* 操纵子中，L4 蛋白正好作用于 mRNA 的起始部位，从而抑制了 *S10* 及其后基因的翻译。蛋白质与 mRNA 的核糖体结合位点结合，从而阻断了核糖体与 mRNA 的结合，因而引起翻译抑制。

r 蛋白操纵子能受 r 蛋白自体控制，也受 rRNA 水平的影响。自体调节物 r 蛋白与 rRNA 上结合位点结合的程度较其与 mRNA 上结合位点结合的程度强。当存在 rRNA 时，新合成的 r 蛋白与 rRNA 结合，开始核糖体装配，此时，没有游离的 r 蛋白可与 r 蛋白的 mRNA 结合，mRNA 继续翻译 r 蛋白。一旦 rRNA 合成减慢或停止，游离 r 蛋白富集，r 蛋白分子与 r 蛋白的 mRNA 结合，阻止其继续翻译 r 蛋白。

这种调控模式实现两个目标。第一，r 蛋白的水平受细胞生长条件影响，通过对 rRNA 水平的控制，细胞实现对核糖体所有成分合成的控制。第二，由这些操纵子编码的其它蛋白质的合成不受 r 蛋白基因翻译的影响。例如，*rif* 操纵中，RNA 聚合酶的 β－亚基不被 L10 中的 r 蛋白控制，而由其自身进行自体调控。由此可以理解，为什么合成 RNA 聚合酶的量较核糖体成分的量少。

4.T4 噬菌体蛋白 p32 合成的自体调控

T4 噬菌体基因 *32* 的基因产物是蛋白 p32。p32 能对其自身的蛋白质合成进行自体

调控。蛋白质 p32 在遗传重组、DNA 修复、DNA 复制等过程中与单链 DNA 结合，发挥重要作用。基因 *32* 无义突变能引起无活性蛋白质的过度表达，因此，当 p32 蛋白功能被抑制时，会产生更多的 p32 蛋白。因为基因 *32* 的 mRNA 很稳定，因此这种调控作用发生在翻译水平上。

图 7－22 示意了基因 *32* 控制过程的模型。当噬菌体感染的细胞内存在单链 DNA 时，单链 DNA 与其结合，使其不能与自身 mRNA 结合。然而，缺乏单链 DNA 并有富余 p32 时，p32 蛋白阻止自身 mRNA 的翻译。通过 p32 与 mRNA 的结合，阻止翻译的起始，可直接介导这种效应，这可发生在核糖体结合位点周围富含 A-T 的区域。要使这种控制环节有效地进行，mRNA 上 p32 结合位点应有两个特点：

1) p32 与基因 *32* mRNA 上结合位点的亲和力应明显低于与单链 DNA 的亲和力。结合 RNA 的平衡常数确实比结合单链 DNA 的低，两者相差约 2 个数量级。
2) p32 与 mRNA 的亲和力应明显高于与其他 RNA 序列的亲和力。这种亲和力与 mRNA 的碱基组成和二级结构有关。基因 *32* mRNA 与 p32 结合的调控区不形成二级结构。

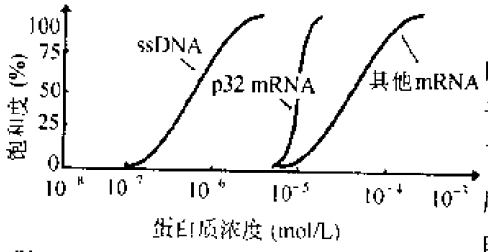


图 7－23 基因 *32* 蛋白结合相关分子的亲和力

用已知的平衡常数，可以绘出 p32 结合靶位点的程度与 p32 蛋白浓度的曲线(图 7－23)。浓度低于 10^{-6} mol/L 时，p32 与单链 DNA 结合。浓度大于 10^{-6} mol/L 时，p32 与基因 *32* mRNA 结合。浓度更大时，p32 则也能与其他 mRNA 序列结合。说明要实现自体调控，p32 的含量应低于 10^{-6} mol/L，也就是每个细菌内约 2000 个分子，这与实际测到的水平相当，实际测到的量是每细胞有 1000～2000 个分子 p32 蛋白。

自体调控的特点是每个调控作用都用专一的，即调控蛋白仅作用于负责指导自身合成的 mRNA。T4 噬菌体基因 *regA* 编码的蛋白则是一个更常见的翻译调节物，它阻抑早期感染期间转录的几个基因的表达。RegA 蛋白通过与 30 S 亚基竞争 mRNA 上的起始位点而阻止这些基因 mRNA 的翻译。其作用正相当于结合多种操纵子的阻抑蛋白的作用。

自体调控是参与构建大分子装配蛋白质合成的常见调控类型。装配好的颗粒本身由于太大、太多，且有严格的空间定位，不适于作为调控物。但是其组成成分的合成可能受某些游离成分的影响。如果由于某种原因装配通路被阻断，则游离成分聚集，装配大分子的成分的合成被关闭。

二、应急调控

细菌处于极差的生存条件下，由于缺乏足够的氨基酸，蛋白质合成受到抑制，因而会关闭大量的代谢过程。这就是应急应答(stringent response)或应急调控(stringent control)。它是细菌为了渡过困难时期而存活下来的适应性表现。细菌仅进行很少的有限的代谢过程以节约使用有限的资源，当营养条件得到改善时，细菌将停止这种应急调控，重新开放各代谢过程。

应急应答使 rRNA 和 tRNA 的合成大幅度下降，下降到原来的十至二十分之一，这使得总 RNA 合成的量仅为应答前的 5%～10%，某些 mRNA 的合成也下降，下降至原有水平的三分之一。蛋白质降解速度加快，核苷酸、糖和脂类等的合成量也下降。

应急应答引起两种异常核苷酸大量增加，一个是腺苷四磷酸 ppGpp，腺苷的 5′和 3′位上都分别连有二磷酸基团；另一个是腺苷五磷酸 pppGpp，腺苷的 5′位上连有三磷酸基团，3′位上连有二磷酸基团。这些核苷酸是典型的小分子效应物，能与目标蛋白结合，使其活性发生改变。有时将它们合称为(p)ppGpp。(p)ppGpp 能协同调控细胞内的很多过程。应急应答可引起(p)ppGpp 大幅度增加。细菌生长速度与(p)ppGpp 水平总体上呈反比关系。

缺乏任何一种氨基酸或引起任何一种氨基酰- tRNA 合成酶失活的突变都能导致应急应答。整个应急应答系列反应的触发器是位于核糖体 A 位点上的未装载氨基酸的 tRNA(空载 tRNA)。正常情况下，由 EF-Tu 将氨基酰- tRNA 放在核糖体 A 位点上，但是当缺乏相应的氨基酰- tRNA 时，空载 tRNA 占据这个空位，核糖体上的蛋白质合成被阻断，引发空转反应(idling reaction)。

通过对松弛型突变体(relaxed mutants，*rel* 突变体)的研究，发现(p)ppGpp 由空转反应产生。*rel* 突变体不表现应急应答，所以，氨基酸的缺乏不会引起稳定 RNA(tRNA 和 rRNA)合成的下降，也不会出现应急应答中常见的其他改变。

最常见的松弛型突变的位点位于 *relA* 基因上，*relA* 基因编码的蛋白称为应急因子(stringent factor)。正常情况下，细胞内这种因子的量很少，每 200 个核糖体才有不足一个分子的这种因子，这种应急因子与核糖体相结合，因此只有很少核糖体能发生应急应答。

如果 A 位点被空载 tRNA 占据，应急细菌的核糖体在体外能合成 ppGpp 和 pppGpp。松弛型突变体的核糖体则不能进行类似的合成，但如果加入应急因子，则能进行合成。

应急因子(RelA)能催化 GTP 与 ATP 反应，生成 pppGpp，催化 GDP 与 ATP 反应，生成 ppGPP，因此根据应急因子的这种催化活性，其正式名称应该是(p)ppGpp 合成酶。与利用 GDP 为底物合成 ppGpp 相比，RelA 酶更多地利用 GTP 为底物进行合成反应，因此 RelA 酶催化合成的主要产物是 pppGpp。然而 pppGpp 能经几种酶转化为 ppGpp。这些酶能进行去磷酸化反应，如翻译因子 EF-Tu 和 EF-G。pppGpp 转化产生 ppGpp 是最常见的途径，在通常情况下，ppGpp 是应急应答的效应物。当细菌的生存条件恢复正常时，就要降解 ppGpp。一种名为 *spoT* 的基因编码降解 ppGpp 的酶。它能以约 20 秒的半衰期快速将 ppGpp 分解为 GDP，因此应急应答能随着(p)ppGpp 的消失很快逆转。*spoT* 突变株能提高 ppGpp 的水平，结果生长变慢。

核糖体的状态会影响 RelA 酶的活性。除了 *relA* 基因突变引起松弛型突变外，在另一基因的突变也可引起松弛型突变，这个基因一开始称为 *relC*，后来发现它就是编码 50 S 亚基蛋白 L11 的基因 *rplK*。L11 位于 A 和 P 位点附近，参与同反密码子配对的空载 tRNA 进入 A 位的反应。此蛋白或其他一些成分构型的改变能激活 RelA 酶，从而使空转反应取代转肽反应。

每一次(p)ppGpp 的合成都引起空载 tRNA 从 A 位上释放出来，因此(p)ppGpp 的合成是空载 tRNA 水平应答的继续。在应急条件下，当空载 tRNA 按密码子配对原则，进入

A 位后，核糖体上进行的正常肽链合成过程暂停。空载 tRNA 的进入，引发(p)ppGpp 分子的合成，(p)ppGpp 又使空载 tRNA 再次释放，然后再次进入，再次引发(p)ppGpp 合成，直到有氨基酰- tRNA 存在，使核糖体能继续进行肽的合成，而结束新一轮空转反应。

ppGpp 是调控一些反应的效应物，能抑制转录。已报道了 ppGpp 的许多作用，其中主要有两个方面：①抑制编码 rRNA 操纵子的启动子的转录起始；②多数或大多数基因转录的延伸过程受 ppGpp 抑制。

三、弱化作用

无论是原核生物还是真核生物，RNA 的二级结构都参与表达调控。其中最常见的调控方式是 RNA 分子在分子内采用不同的碱基配对方式，形成不同的二级结构而参与调控。当不同的结构对是否允许终止存在差异时，改变二级结构可对转录终止进行调控。另一种控制构型的方法是对 RNA 切割，切除一个 RNA 片段后，引起剩余部分的构型发生变化。调控的方法还可能是一个小 RNA 分子与目标 RNA 分子直接形成碱基配对，从而控制目标 RNA 分子的表达活性。

一些操纵子进行弱化调控。弱化作用(attenuation)是控制 RNA 聚合酶通读弱化子能力的调控机制。弱化子(attenuator)是位于转录单位开始区的一种内部终止子。弱化作用的共同特点是一些外部事件控制内部终止所需的发夹结构的形成。如果形成发夹结构，则发生阻止 RNA 聚合酶转录结构基因的终止作用。如果不形成发夹结构，RNA 聚合酶通过终止子，基因得以表达。控制弱化作用的二级结构的改变由核糖体在 mRNA 上的位置所决定。图 7－24 说明终止需要核糖体能翻译 mRNA 的一个前导片段。当核糖体能翻译前导区时，在终止子 1 上形成终止发夹结构。但是当核糖体还能翻译前导区时，发夹结构在另外位置上形成，结果终止区保持单链，RNA 聚合酶转录编码区。因此，外部环境对核糖体在前导区移动影响的能力控制着这种抗终止作用。

弱化作用是在色氨酸操纵子中发现的。色氨酸操纵子有 5 个连续排列的结构基因，它们编码使分支酸(chorismic acid)转化为色氨酸的三种酶。这些基因的转录从它们的上游共同的启动子开始(图 7－25)。与启动子相邻的操纵基因能与另一位点上基因 *trpR* 编码的阻抑蛋白结合。在色氨酸操纵子操纵基因与第一个基因的编码区之间有一前导区。结构基因的转录部分终止在内部终止子 *trpt* 上，终止子 *trpt* 位于编码区的 3′端外侧，大约 36 bp 处。其后大约 250 bp 处，有 ρ 因子决定的终止子 *trpt′*，*trpt′* 对操纵子的功能来说，是非必需的。同样类型的操纵子也存于大肠杆菌和鼠伤寒沙门氏菌(*Salmonella typhimurium*)中。

除启动子-操纵基因复合物外，还有一个位点对色氨酸操纵子起调控作用。这个位点是因删除操纵基因与 *trpE* 编码区间的序列，结构基因的表达增强而发现的。这种作用与阻抑作用无关，删除后，基础水平转录和去阻抑转录都增强。因此，此位点影响 RNA 聚合酶从启动子出发后发生的转录过程。启动子和 *trpE* 基因间的序列有两个重要特点：

1) 它是能编码含 14 个氨基酸的前导肽的短的编码序列。

2) 弱化子(内部终止子)能阻碍结构基因的转录。RNA 聚合酶在此处终止，形成一

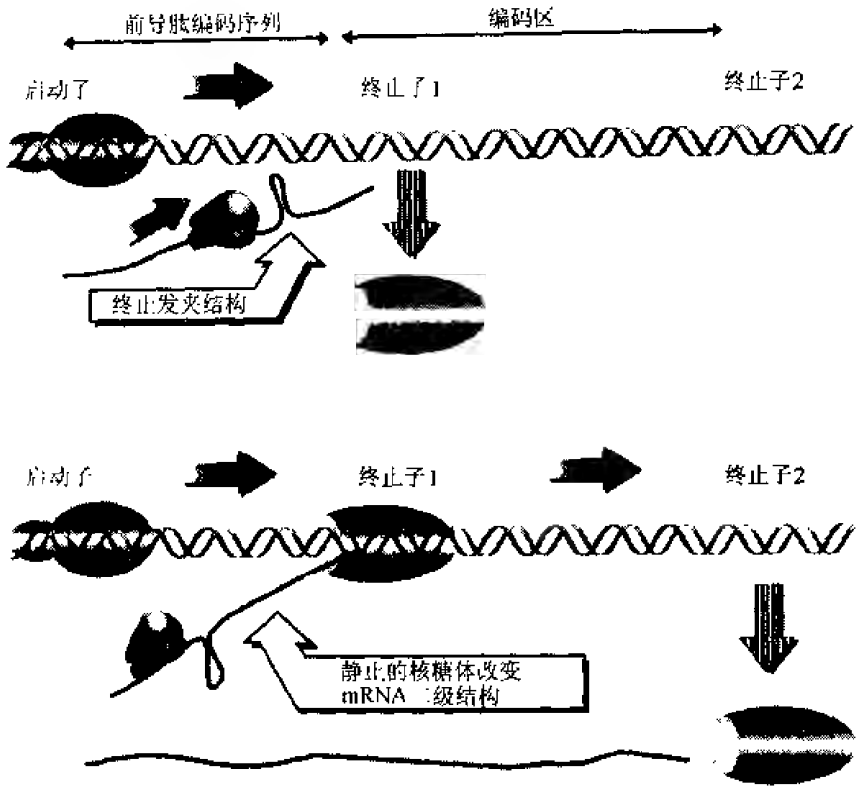


图 7-24 通过核糖体移动改变 RNA 二级结构对终止作用的控制

个 140 bp 的转录物。

弱化子上的终止受色氨酸含量的影响。色氨酸含量高时,弱化子上能终止。但是,缺乏色氨酸时,不能引起终止,RNA 聚合酶继续翻译结构基因。

阻抑作用和弱化作用应答色氨酸含量变化的结果相同。当色氨酸存在时,操纵子被阻抑,大多数 RNA 聚合酶脱离启动子,并在弱化子上终止。不含色氨酸时,RNA 聚合酶能与启动子结合,在弱化子上不发生转录终止。

当色氨酸存在时,终止有效,弱化子使得只有约 10% 的 RNA 聚合酶能继续参与转录。在缺乏色氨酸时,所有的 RNA 聚合酶都能经过弱化子继续参与转录。加上取消阻抑增加的约 70 倍的表达,缺乏色氨酸时色氨酸操纵子表达量增强约 700 倍。

弱化子转录终止是怎样对色氨酸含量做出应答的呢? 前导区序列可能与此有关。从图 7-25 可见,前导区序列含一个核糖体结合位点,其 AUG 密码子后有一短编码序列,它含有连续的色氨酸密码子。当细胞中色氨酸用尽时,核糖体开始翻译前导肽,但是当它们遇到色氨酸密码子时停下来。mRNA 的这种序列说明这种核糖体的暂停会影响弱化子上的终止。

前导序列能形成两种碱基配对结构,核糖体经过前导区继续转录的能力控制着这两种结构的转换,这种结构决定 mRNA 是否形成终止所需的结构。图 7-26 列出了这些结构。第一个结构中,1 区与 2 区配对,3 区与 4 区配对,3 区与 4 区配对形成发夹结构,接着发夹结构就是 U_8 序列,U_8 是内部终止子的基本信号。

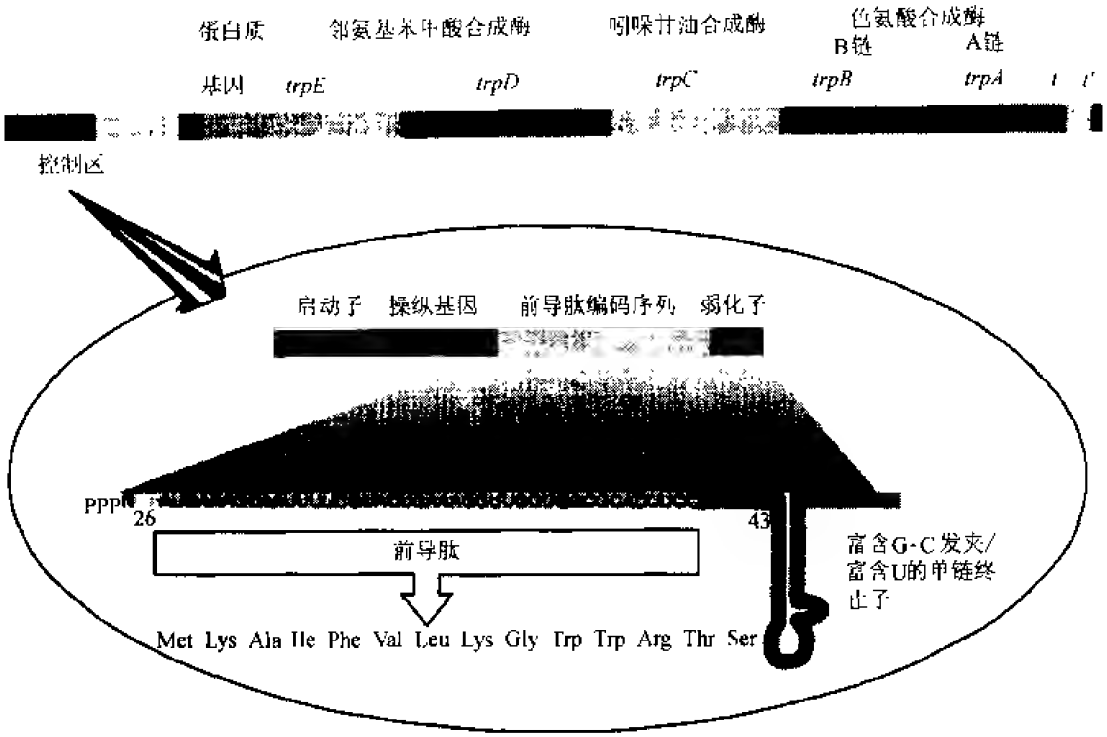


图 7 25 色氨酸操纵子

如果 1 区不能与 2 区配对，则形成另一种结构。此时，2 区能与 3 区配对，4 区则没法配对，因此只能保持单链，不能形成翻译终止所需的发夹结构。

在细菌中，转录进行到前导区第 90 碱基处时，转录暂停，直到核糖体翻译前导肽。然

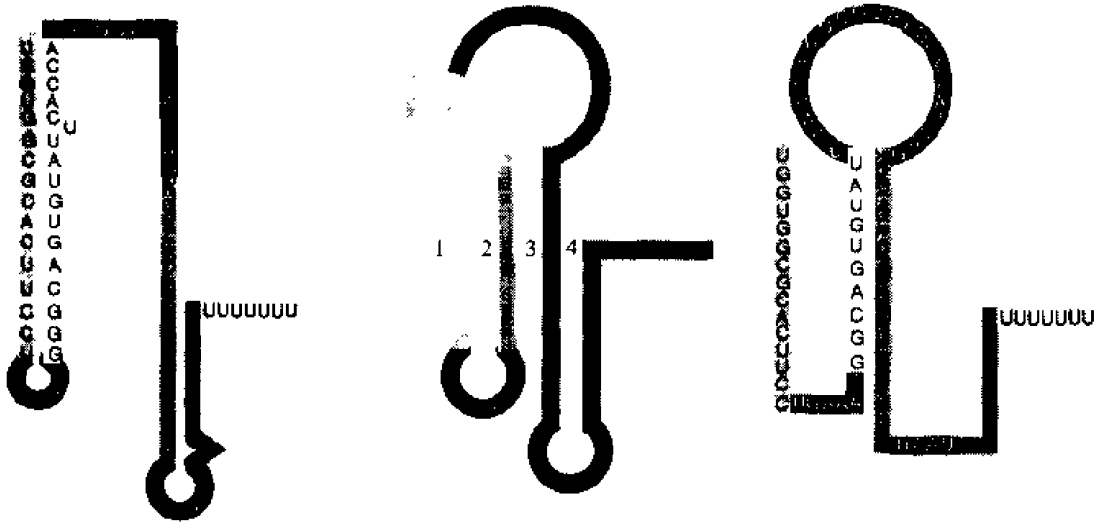


图 7-26 *trp* 前导区的两种配对构型

2 区可与 1 区和 3 区配对。3 区可与 2 区和 4 区配对。左图中，1 区和 2 区配对，3 区与 4 区配对，形成终止子发夹结构。右图中，2 区和 3 区配对，终止子区为单链结构，终止所需的发夹结构不能形成。

后 RNA 聚合酶释放出来并移向弱化子位点。当它到达弱化子区时，决定是否继续转录的二级结构已经形成。这是因为色氨酸操纵子前导肽中含有色氨酸密码子，当细胞中有色氨酸存在时，核糖体能够顺利地翻译出整个前导肽而在终止密码子 UGA 处停下来。这时核糖体占据了 1 区和 2 区的一部分，使 2 区和 3 区不能有效的配对，因而 3 区和 4 区配对产生终止子的发夹结构，于是实现转录的终止。当细胞内色氨酸缺乏时，核糖体停顿在两个色氨酸密码子上。这时，核糖体占据了 1 区，而留下完整的 2 区以便与转录出的或即将转录出来的 3 区形成二级结构，这样当 4 区转录出来后仍然是单链状态，不能形成带有发夹结构的终止子于是转录继续进行下去。

细菌利用弱化作用控制操纵子的现象比较普遍。至少有 6 个编码与氨基酸合成有关的酶的操纵子采用这种控制机制。蛋白质合成中氨基酸的水平反馈控制用于该氨基酸合成的酶量。

四、小 RNA 分子调节翻译

到目前为止，我们前面讨论的反式作用调节因子都是蛋白质。但 RNA 也可作为调节因子参与调控，现在已经有几个小 RNA 分子调节基因表达的例子。与蛋白质调节物一样，RNA 合成好后，扩散到含特异核苷酸序列的靶位点。RNA 调节物的靶分子是一个单链核酸序列。RNA 调节物与目标分子互补，形成双链区而发挥作用。调节物 RNA 作用机制有两种，一种是 RNA 分子与靶分子形成双螺旋，通过形成和封闭特异位点直接影响靶分子的功能；另一种是调节物 RNA 分子与靶分子的部分区域形成双螺旋，从而改变靶分子其他区域的构型，间接影响靶分子的功能。两种调控机制的共同点是都要通过改变靶分子的二级结构来控制靶分子的活性。

RNA 调节物与阻抑操纵子的蛋白质的不同点是 RNA 没有变构的性质，它不能受其他小分子影响改变识别靶分子的能力。因此它的调节作用随着其产生而形成，随着其消失而失去。所以从产生的角度看主要是对其基因转录进行控制，从分解的角度主要是由催化其降解的酶来控制。

反义 RNA 就是一种 RNA 调节物，在细菌细胞内作为调节物进行一些调节。反义 RNA 有三种调节方式。反义 RNA 与 mRNA 上核糖体结合位点结合，使核糖体不能结合，使得翻译不能起始。反义 RNA 也可与 RNA 结合，形成双螺旋结构，由于所形成的双螺旋结构成为内切酶的特异底物，使与其结合的 RNA 变得不稳定。反义 RNA 也可与转录物结合，使得转录提前终止。

小 RNA 的合成直接控制大肠杆菌 *ompF* 基因的翻译(图 7－27)。编码大肠杆菌两种外膜蛋白质的基因 *ompC* 和 *ompF* 的表达受培养基的渗透压控制，这两个基因参与调控对渗透压变化的应答。基因座 *envZ* 编码一种作为渗透压感受器的受体蛋白，渗透压增加时，EnvZ 蛋白激活另一种蛋白质—基因 *ompR* 的产物，OmpR 蛋白是激活两种基因转录的一种正调节物，它所激活的两种基因是中心调节区的结构基因 *ompC*(图 7－26 中未画出)和调节基因 *micF*。*micF* 的产物是一个 174 碱基大小的 RNA 分子，它是典型的反义 RNA，反义 RNA 是指能与 mRNA 互补结合，从而阻断 mRNA 的翻译的 RNA 分子。

渗透压增加引起 MicF RNA 合成，MicF RNA 关闭 *ompF* mRNA 的翻译。MicF

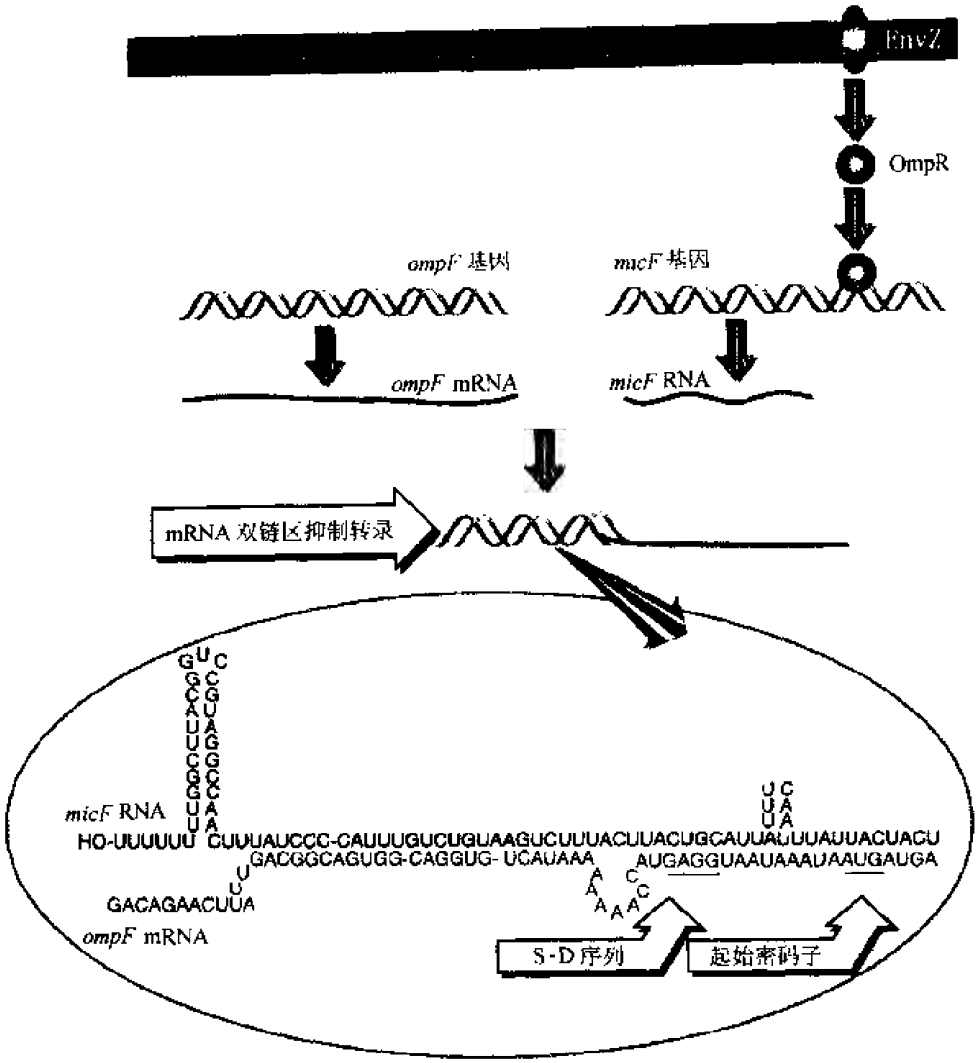


图 7-27 *micF*RNA 对 *ompF* mRNA 翻译的调控

渗透压增压激活 EnvZ，EnvZ 激活 OmpR，OmpR 诱导 *micF* 和 *ompC*（图中未显示）转录，*micF*RNA 与 *ompF* RNA 的 5′区互补，两者结合后阻止 *ompF* 的翻译。

RNA 与 OmpF mRNA 分子中含有翻译起始的核糖体结合位点的区域互补，因此 MicF RNA 能发挥调节物的功能，与 OmpF mRNA 结合阻止其翻译。它也能与 OmpF mRNA 形成双螺旋结构区，使 OmpF mRNA 不稳定，例如使它对水解双链的核糖核酸酶敏感就能使 OmpF mRNA 变得不稳定。

秀丽隐杆线虫中也存在一个编码 RNA 的调节基因，基因产物 RNA 可与一 mRNA 互补。这种调节基因是 *lin4*，它所调节的对象是 *lin14* 基因，Lin14 参与幼虫发育的调控。基因 *lin4* 编码两个大小分别为 22 核苷酸和 61 核苷酸的小 RNA，两者的终止点不同。*lin4* 的转录产物与一个 10 碱基的序列互补，该序列在 *lin14* 的 mRNA 非翻译区的 3′端重复 7 次。现在仍不清楚 *lin4* RNA 怎样关闭 *lin14* 的表达。一种可能的机制是双链的形成会阻止 *lin14* mRNA 从细胞核出来，到细胞浆中去。另一种可能的机制双链的形成

会干扰翻译过程。此调节系统还有一个特别之处，就是调控位点在基因的 3′端。

反义 RNA 的合成能抑制原核生物和真核生物的 mRNA 的翻译。利用这个原理已将人工合成的基因引入生物体内，使其转录出反义 RNA，达到抑制与反义 RNA 互补的目的 mRNA 翻译的目的。

（欧阳红生）

参 考 文 献

[1] Miller J et al. The Operon. New York: Cold Spring Harbor Laboratory. 1978. 11
[2] Lewin B. Genes. New York: Oxford University Press and Cell Press. 1997. 335
[3] Ochler S. EMBO J, 1990. 9: 973
[4] Friedman A M et al. Science, 268: 1721
[5] Lewis M et al. Science, 1996. 271: 1247
[6] Botsford J L et al. Microbiol Rev, 1992. 56: 100
[7] Bram R J. Cell, 1980. 19: 393
[8] Gold L. Ann Rev Biochem, 1988. 57
[9] Izant J G et al. Cell, 1984. 36: 1007

第八章　真核生物基因表达调控

第一节　概　　述

各种真核细胞表型的差异主要是其编码蛋白质基因表达不同引起的。这些编码蛋白质的基因的表达都经 RNA 聚合酶 II 转录。理论上，这些基因的表达可能都在数个阶段中的某一个阶段受到调控。因此，“控制水平”的概念是指基因的表达并非一旦开始，就必然自动进行至最后完成。基因的表达能在几个连续步骤中的任一步中以基因特有的方式受到调控。我们至少可以分出 5 个控制点，即：

基因结构的激活（活化）→转录起始 →转录过程 →胞浆转运 →mRNA 翻译

第一步的存在说明基因存在两种结构状态。与基因组中的大多数基因状态相比，细胞中处于“活化”状态的基因得到表达。这种结构的改变与转录的作用不同，表明此基因可以转录。因此，变成“活化”结构是基因表达的第一步。

处于活化状态基因的转录由转录起始阶段控制，即通过 RNA 聚合酶与启动子相互作用进行控制。这种控制可在体外实验系统中进行分析。对大多数基因来说，转录起始的控制是最主要的控制点，也就是说最常见的调控水平是转录起始水平的调控。

在随后的转录过程中的调控，例如通过抗终止作用进行的调控，在真核细胞中是否存在目前仍没有确切证据。

初级转录物经 5′端加帽及 3′端加多聚腺苷酸尾等加工，割裂基因转录物中的内含子还要经剪接去掉。成熟的 RNA 必须从胞核出到胞浆中。在核 RNA 水平上，还可通过改变剪接类型实现调控蛋白质产物的类型。

最后，胞浆中一个特定的 mRNA 是否被翻译仍被调控。在成年体细胞中，采用这种调控机制的例子还很少，但在胚胎时，确实有这种调控的例子，即一些特异蛋白质因子阻断某些 mRNA 翻译的起始。

第二节　真核生物基因转录调控[1~7]

绝大多数调控发生在转录起始阶段，但由于基因表达的控制可发生在多个阶段，RNA 产物的产生并不是就必定会形成蛋白质产物。组织特异性基因表达调控是真核细胞分化的核心，控制胚胎发育的转录因子中也可看到这方面的例子。一个转录因子可以对许多目的基因进行调控。我们试图找到关于调控模式的两个问题的答案，即什么决定了转录因子的目的基因？转录因子的活性怎样受到内外信号的调控？

一、基因转录的顺式调控元件

真核基因的顺式作用元件按照功能可以分为启动子、增强子以及沉默基因（silencer）。

(一) 启动子的选择

RNA 聚合酶 II 的启动子有含 TATA 框的典型启动子和不含 TATA 框的非典型启动子两种。

1. TATA 框启动子

有 TATA 框的典型启动子是上游启动子和增强子产生诱导效应所必需的。有时一个基因上有串联着的两个 TATA 框,它们可分别地或有侧重地对不同的诱导物作出应答;在某些情况下,也参与组织特异性的选择。α 淀粉酶基因在唾液腺和肝脏两种组织中分别选择了相距 2.8 kb、转录效率不同的两个转录起始点,以保证唾液中酶的活性远高于肝脏组织。

2. 不典型的启动子

有少数基因没有典型的 TATA 框启动子序列,有的无 TATA 框启动子的序列富含 GC,即具有 GC 框,有的则没有 GC 框。

(1)*富含 GC、无 TATA 框的基因转录*　不依赖或缺少典型启动子结构的基因转录起始是不规则的,并且只有基础水平表达;如小鼠的金属硫蛋白基因在睾丸组织中低表达,转录起始点散布在 −160 bp～+1 之间,其中 60% 的转录起始于 −50 bp 到 −150 bp 的区间内。

持家基因(housekeeping gene)是以维持细胞正常结构、基本生命活动所需的蛋白质的编码基因。这些基因的 5′上游有时没有 TATA 框,只有富含 GC 的上游序列,其中常有一个以上的 SP1 结合位点(GGGCGG),SP1 普遍存在于多种细胞中对基础转录的活化有重要作用。持家基因常有多个转录起始点,分布跨度也较大,具有这类典型散在起始点的基因有二氢叶酸还原酶和人表皮生长因子受体的基因等。另外一些组成型表达的基因如组蛋白 H2B 等,在其近侧启动子区有 Oct1 结合的 8 个核苷酸序列存在。若将 H2B 启动子区中的 Oct 序列删除,则 H2B 转录不再具有细胞周期特异性。可见基因的启动子元件与转录起始方式和表达性质有密切关系。

(2)*转录起始子与无 TATA 框、GC 框的基因转录*　TATA 框与转录起始子(initiator, Inr)是决定转录起始(+1)位点的关键序列。有些基因虽有 TATA 框但属于较弱的启动子,此时 RNA 聚合酶 II 的转录起始于一个或数个紧密成簇的起始元件 Inr 的 A 位上。这种元件的保守性序列(5′- PyPyCAPyPyPyPyPy - 3′)很有限。这类基因有果蝇发育时的同源转换基因 *ubx* 和 *antp*、T 淋巴细胞特异性的 T 细胞抗原受体(TcR)β 链编码基因、原癌基因 *lck* 等。具有淋巴细胞分化特异性表达的末端脱氧核糖核苷转移酶(TdT)启动子上没有 TATA 框,却在其转录起始位点附近 −6～+11 的 17 个核苷酸(5′-GCC CTC ATT CTG GAG AC - 3′)处形成起始子,控制着 *TdT* 基因在前 B 和前 T 淋巴细胞中特异地由 +1 位的 A 起始准确转录。若在 Inr 上游加入 TATA 框或上游启动子如 GC 框等都可明显提高 Inr 的转录效率。有趣的是在腺病毒主要晚期启动子中有典型的 TATA 框却也有 Inr、其 −6～+5 范围内的序列与 *TdT* 的 Inr 完全相同,提示在上述典型

的启动子中，TATA 框可与 Inr 同时起作用。

（二）增　强　子

在真核生物基因组很长距离上都可能发生基因表达的调控。一些真核生物的基因受增强子（enhancer）序列调控。增强子是真核细胞中通过启动子来增强转录的一种远端遗传性控制元件。有效的增强子可以位于基因的 5′端，也可位于 3′端，还可位于基因内的内含子区。增强子区在真核细胞中一般跨度为 100～200 bp。增强子和启动子同样是由多个独立的、具有特征性的核苷酸序列所组成的。其基本的“核心”元件（element）常由 8～12 bp 组成；可以有完整的或部分的回文结构，并以单拷贝或多拷贝串联的形式存在。一般认为各个独立序列彼此间隔 50 bp 以内，至少 2～3 个同时存在，就能有效地协同促进转录活性。

1. 增强子的特性

增强子首先发现于 SV40 病毒中。早期启动子 5′上游约 200 bp 的 DNA 片段连接在其他基因旁侧可促使该基因转录效率提高 100 倍，该片段可远离基因转录起始点达 3 000 bp 以上。SV40 病毒基因增强子区有前后两个 72 bp 的重复序列，其中的“核心”元件是 5′-GGTGTGGAAAG-3′。随后在免疫球蛋白重链 J 与 C 基因间的大内含子中发现了第一个 μ 基因的细胞增强子。表明真核细胞中也有远距离调控的增强子存在；并发现这种激活基因转录的细胞增强子具有细胞特异性；该增强子在免疫球蛋白基因的“内部”，即位子转录起始的“帽”位点下游。

增强子的特性可归纳为 7 个方面。①增强子能（通过启动子）提高同一条 DNA 链上靶基因转录的速率；②增强子对同源或异源基因同样有效；③增强子的位置可在基因 5′上游、基因内或基因的 3′下游序列中；④增强子在 DNA 双链中没有 5′与 3′固定的方向性；⑤增强子可远离转录起始点，通常在 1～4 kb（个别情况下可远离转录起始位点达 30 kb）起作用；⑥增强子一般具有组织或细胞特异性；⑦增强子的活性与其在 DNA 双螺旋结构中的空间方向性有关。

2. 增强子的作用机制

由于增强子的活性与半周 DNA 双螺旋（5 bp）的奇、偶位点有关，且常在电镜下呈现球状结构，提示增强子作用受 DNA 双螺旋空间构象影响。增强子在转录起始点远端起作用可能有 3 种方式。①增强子可以影响模板附近的 DNA 双螺旋结构，如导致 DNA 双螺旋弯折，或在反式因子的参与下，以蛋白质之间的相互作用为媒介形成增强子与启动子之间“成环”连接的模式活化转录；②将模板固定在细胞核内特定位置，如连接在核基因上，有利于 DNA 拓扑异构酶改变 DNA 双螺旋结构的张力，促进 RNA 聚合酶 II 在 RNA 链上的结合和滑动；③增强子区可以作为反式作用因子或 RNA 聚合酶 II 进入染色质结构的“入口”。

3. 增强子的种类

(1)细胞特异性增强子　　许多增强子的活性虽然只在个别的细胞或组织中出现,然而其中绝大部分组织特异性的增强效应都是由于不同细胞(或组织)中含有特异的 DNA 调控蛋白的活性(反式调控机制)所决定的。病毒增强子对于细胞的选择性和细胞中固有基因增强子的组织特异性是有区别的。如胰腺内分泌性的 β 细胞中胰岛素基因 5′上游 −103～−353 bp 是强的组织特异性增强子。

(2)诱导性增强子　　诱导性增强子的活性通常需要有典型的启动子参与。金属硫蛋白基因可在多种组织中转录,又可受类固醇激素、重金属和生长因子等多种因素诱导。其调控元件除启动子外,也相应地包括两个基础水平的元件以及诱导性类固醇激素应答元件金属应答元件等。基础水平元件有助于基因的组成性表达,两个诱导性元件可被相应因素诱导而促进转录。

(三)沉默基因

沉默基因是一种负调控元件,参与基因表达的负调控。T 淋巴细胞可分为两种不同的 T 抗原受体 α/β、γ/δ。由于编码 α 和 δ 的基因同在一个基因座位中而 α 链恒定区 *Cα* 基因 3′下游的 α 增强子在各种 T 细胞均有潜在活性,因此在 T 细胞分化时需要通过各种正负调控机制进行精密的选择,才能正确表达。已知的负调控元件一个紧接在 *Cα* 基因的 3′下游,另一个在 *Cα* 增强子的上游。对 α 特异的沉默基因可阻止 *Jα* 基因在 γ/δ 型细胞中参与重排,而相应的 δ 增强子将促进 δ 基因重排到位。可见 α 沉默基因的 *TCR* 基因的 α/δ 基因座转录和重排中起着重要作用。这些负调控元件可不受距离和方向的限制,并可对异源基因的表达起促进作用。

(四)转座元件

转座元件(transposable element)的基因调控可能是由于转座元件的插入,带来新的序列特异性 DNA 结合蛋白及其结合位点,这些序列可以增强子的方式远距离地调控基因转录;当这样的调控序列转座到原癌基因附近就可使之转变为癌基因,进而产生出肿瘤细胞。由于真核细胞基因组中蛋白质编码区仅占 2%,因此转座元件的移动很少引起外显子的改变,而主要是通过其调控作用使机体在长期的进化过程中,适应环境压力选择出更多的异质品系。

二、通用调控中的调控效应元件

受共同控制的一组基因共用一个受转录调控因子识别的启动子元件。使基因应答此类因子的元件被称为效应元件(response element),如热激效应元件(heat shock response element, HSE)、糖皮质激素效应元件(glucocorticoid response element, GRE)和血清效应元件(serum response element,SRE)。

表 8-1 一些转录因子及其所识别的元件

调节剂	效应元件	保守序列	因子
热激	HSE	CNNGAANNTCCNNG	HSTF
糖皮质激素	GRE	TGGTACAAATGTTCT	受体
佛波酯(phorbol ester)	TRE	TGACTCA	API
血清	SRE	CCATATTAGG	SRF

效应元件有启动子上游元件或增强子相同的特点。它们含短的保守序列,不同基因中的效应元件拷贝数很接近,但不完全相同。调控因子的结合区在保守序列上,保守序列两端覆盖距离不长。在启动子上,效应元件距转录起始点的距离并不固定,但通常在转录起始点的上游 200 bp 内。只要单个效应元件就可以受调控因子的调控,但有时,有多个拷贝的效应元件出现。

效应元件可能位于启动子内,也可能位于增强子内。典型的元件只存在于启动子或增强子二者之一内,通常,HSE 存在于启动子中,而 GRE 存在于增强子中。我们认为所有元件都以同样的一般原理发挥作用。即基因受识别启动子或增强子中序列的特定蛋白质调控,这种特定蛋白质起转录因子的作用,是 RNA 聚合酶启动转录所必需的。活化蛋白仅在基因被表达这个特定条件下可以得到;它的缺乏意味着启动子就不能按这种特定程序激活。

热激应答中多个基因由单一的转录因子控制就是这种情况。热激应答在原核和真核生物中许多基因的表达控制中很普遍,温度增加关闭一些基因的转录,同时开放热激基因(heat shock gene)的转录,从而引起 mRNA 翻译的变化。原核生物和真核生物热激基因控制的模式不同。在细菌,合成一种新的 σ 因子,使 RNA 聚合酶全酶识别热激基因启动子中的 -10 位上的共同序列。在真核生物中,热激基因也含共同保守序列(HSE),这些保守顺序离转录起始位置的距离不固定,但都能被专一的转录因子 HSTF 所识别。

黑腹果蝇(*Drosophila melanogaster*)的所有的热激基因都含多拷贝 HSE。HSTF 结合到邻近的应答元件上。HSE 和 HSTF 在进化上保守,黑腹果蝇的热激基因既能在哺乳动物中激活,也能在海胆中激活。果蝇的 HSTF 蛋白与酵母的相似,对含 HSE 序列的 DNA 中,表现出相同的足迹图谱。当细胞热激时,酵母的 HSTF 发生磷酸化,这种修饰使 HSTF 蛋白质激活。

金属硫蛋白(metallothionein, MT)基因提供了一个单一基因是怎样受多种不同的调控机制调控的实例。金属硫蛋白通过与金属结合然后使其排出细胞从而保护细胞免受高浓度重金属的侵害。通常金属硫蛋白基因在基础水平表达,但是当受到重金属离子(如镉)或糖皮质激素诱导时,表达水平提高。其调控区含有数种调控元件,表明其调控的原理是当启动子受一种以种方式调控时,每一调控过程都有特定序列与相应的蛋白质结合。

MT 基因启动子结构成分见图 8-1。此图中主要特点是可激活转录的元件密度很大。两个组成型启动子元件(constitutive promoter element)是 TATA 框和 GC 框,位置变化不大,紧靠着转录起始点,基础水平的组成型表达(constitutive expression)也需要基础水平元件(basal level element,BLE)。BLE 属于增强子范畴。尽管 BLE 位于转录起始点的附近,但移到其他位置,仍起作用。在其他增强子中有 BLE 的相关序列,这些序列可与

SV40 增强子结合蛋白质结合。

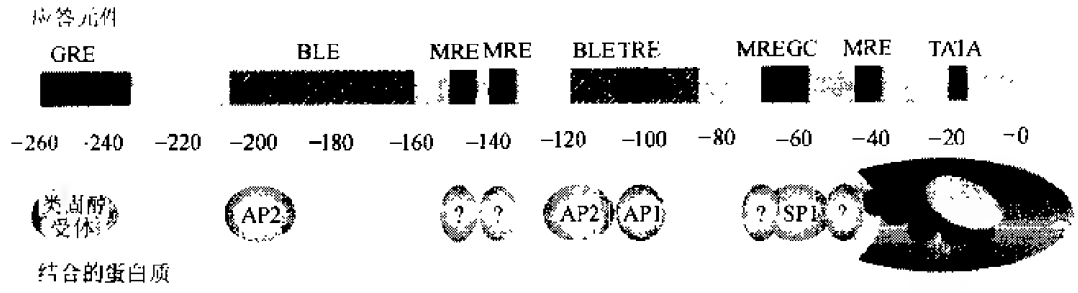


图 8-1 启动子和增强子中含组个"组成型元件的人金属硫蛋白的调控区
启动子有金属诱导元件,有应答糖皮质激素的增强子元件。图上部的缩写为调控元件,下部的缩写为主调控元件结合的蛋白质。

TRE 含一些增强子的共有序列,包括一个金属硫蛋白的 BLE 和 SV40 病毒的 72 bp 重复序列。TRE 上有转录因子 AP1 的结合位点,两者相互作用是引起组成型表达的机制之一。然而,AP1 的结合也有其第二功能。TRE 负责应答佛波酯如 TPA(启动肿瘤的一种试剂)等信号,也通过 AP1 与 TRE 相互作用实现。

多份 MRE 序列负责对金属刺激的应答,它们起启动子元件的作用。一价 MRE 就有了对重金属信号应答的能力,多份 MRE 元件则可完成更高水平的诱导。

GRE 负责应答类固醇激素信号。GRE 位子转录起始点上游 250 bp,起增强子作用。GRE 的缺失不影响基因水平的表达,也不影响金属离子诱导的表达。但是应答类固醇激素信号则绝对必要。

金属硫蛋白的调控说明了调控的通用原理,几个不同的元件中的任一个,无论位于启动子中还是增强子中都能独立激活基因表达。缺乏某一激活模式的元件并不影响其他模式的激活。各种元件独立作用,它们之间的排列没有特别严的限制,表明一个与任一元件结合的因子能够通过基础转录装置独立提高转录起始的效率,也许借蛋白质与蛋白质相互作用,使起始复合物更稳定或促进起始复合物的形成。

三、DNA 结合结构域

对各种转录因子的序列进行比较,可发现其基序(motif)的共同点是都与 DNA 结合。基序通常很短,仅为蛋白质结构中的一小部分。在蛋白质与转录装置相互作用中,负责激活转录的基序可以被鉴定出来。用特定的基序与 DNA 结合,得到几组调节转录的蛋白质基团的详细信息。

1. 类固醇受体

类固醇受体(steroid receptor)是一组功能相关的蛋白质,每个受体都经与特定的类固醇结合而得到激活。糖皮质激素受体是分析得最彻底的类固醇受体。类固醇受体与其他受体如甲状腺激素受体和视黄酸受体一起,组成作用方式相同的转录因子超家族。

2. 锌指结构

锌指结构(zinc finger)基序含一个 DNA 结合域。最早发现的锌指结构基序被 TFIIA 因子识别,而 RNA 聚合酶 III 转录 5 S rRNA 基因需要这种 TFIIA 因子。在另一些转录因子中也发现有锌指结构。在类固醇受体中也发现了一种特别形式的这种基序。

3. 螺旋-转角-螺旋

螺旋 转角 螺旋(helix-turn-helix)基序作为 DNA 结合域最早发现于噬菌体的阻抑蛋白中。一个 α-螺旋位于 DNA 的宽沟中;另一个位于穿过 DNA 的角中。在同源域(homeodomain)中有基序的另一种相关形式存在。同源域是首次在与果蝇发育有关的基因编码的几个蛋白质中定性的一个序列。这种结构也存在于哺乳动物转录因子中。

4. 螺旋-环-螺旋

两亲性螺旋-环-螺旋(helix-loop-helix,HLH)基序是从一些发育调节物及真核 DNA 结合蛋白中发现的。每种两亲性螺旋一侧有疏水残基面,另一侧有带电残基构成的亲水面。相连的环长度有多种,为 12～28 个氨基酸。此基序使蛋白质形成二聚体,与 DNA 接触的基序附近有一碱性区。

5. 亮氨酸拉链

亮氨酸拉链(leucine zipper)由伸展着的氨基酸组成,氨基酸中每隔 6 个氨基酸就有一个亮氨酸残基。一条多肽链的亮氨酸拉链与另一多肽链上的亮氨酸相互作用形成二聚体。每个拉链相邻部位是展开的正电荷残基,它们与 DNA 结合。

可诱导的转录因子的活性可受一种或几种方式调控。下面简介这些调控方式。

(1)合成转录因子　这类因子是组织特异性因子,它仅在特定细胞中合成。这是典型的调节发育的因子,如同源域蛋白。

(2)修饰　一种因子的活性可通过修饰直接控制。如 HSTF 经磷酸化转变成活性形式。AP1(亚基 Jun 和亚基 Fos 构成的二聚体)经 Jun 亚基磷酸化转变成活性形式。

(3)结合配基　一种因子的活性可通过结合配基而进行激活或抑制。类固醇受体是这种控制的一个例子。配基结合可能影响蛋白质的定位(引起从细胞浆到细胞核的转运),也改变其结合 DNA 的能力。

(4)切割释放出转录因子　当一个蛋白质与核膜或内质网结合时,产生一种转录因子。固醇如胆固醇缺乏引起蛋白质的胞浆区被切割,然后转运到细胞核中,形成转录因子的活性形式。

(5)抑制剂释放　一种因子的有效性可能变化,如因子 NF-κB(在 B 细胞中它能激活免疫球蛋白 κ 基因)存在于很多类型的细胞中。但是它被抑制剂 I-κB 限制在细胞质内。在 B 淋巴细胞中,NF-κB 从 I-κB 中释放出来移到细胞核内激活转录。

(6)与不同伴侣形成二聚体　一个二聚体因子可能有两个伴侣分子。无活性的伴侣可能引起它失活,活性伴侣分子的合成可取代无活性的伴侣分子,使它激活。这种情况可见于 HLH 蛋白中。

下面将更详细地讨论这些蛋白质中的一些与 DNA 的结合及激活作用。通常转录因子与 DNA 双螺旋的大沟内的碱基或磷酸接触，通过 α-螺旋的一个短区域与 DNA 结合。

四、锌指基序———一种 DNA 结合结构域

锌指结构是由图 8-2 所示的结构而得名。锌指结构中，保守氨基酸的小基团与锌离子结合形成蛋白质中相对独立的结构域。两类 DNA 结合蛋白质——传统的“锌指”蛋白和类固醇受体有这种结构。典型的“锌指蛋白”如图 8-2 描述的一样，有一系列锌指结构。一个锌指结构的保守序列是：

Cys-X_2-Cys-X_3-Phe-X_5-Leu-X_2-His-X_2-His

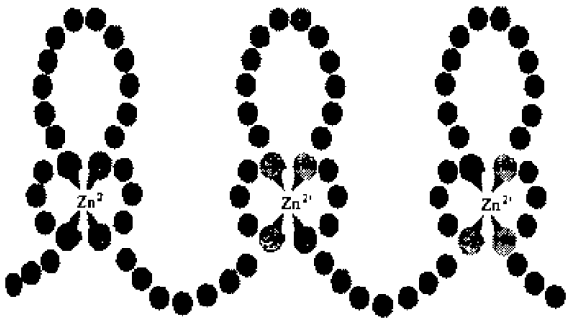


图 8-2 转录因子 SP1 中的锌指结构
转录因子 SP1 有三个锌指结构，每一个都以半胱氨酸及组氨酸残基构成锌结合位点。

其基序根据锌结合位点凸出的氨基酸环命名，如 Cys_2/His_2 锌指。锌位于保守的 Cys 和 His 残基形成的四面体结构中。锌指本身含约 23 个氨基酸，锌指间通常有 7～8 个氨基酸。

锌指是 DNA 结合蛋白的一个通用基序。锌指结构通常串联重复排列在一起，通常都有多个锌指存在。锌指数多少不等，有的有 9 个，几乎占满全部蛋白质（如 TFIIIA 中），有的只有 2 个，仅占一个小结构域（如果蝇的调节物 ADR1）。通用转录因子 SP1 有一个 DNA 结合域，含 3 个锌指。

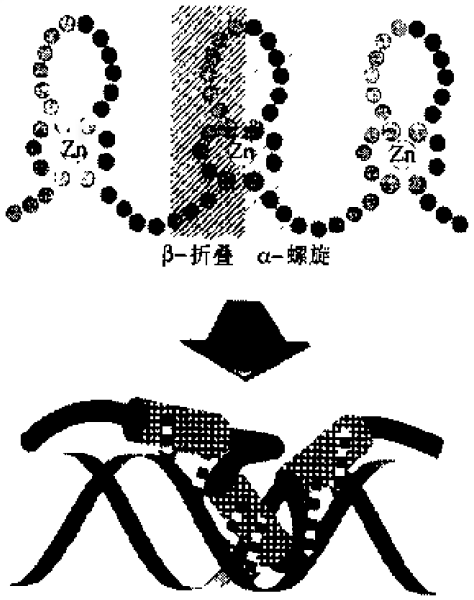


图 8-3 锌指结构蛋白质与 DNA 结合的晶体结构
锌指结构可形成插入 DNA 双螺旋大沟的 α-螺旋，另一侧与 β-折叠相连。

图 8-3 描述了结合含三个锌指结构蛋白质的 DNA 的晶体结构。每个锌指的 C 端部分形成 α-螺旋，与 DNA 结合。N 端形成 β-折叠。β-折叠和锌离子在图中未显示出来，它应位于该图的下方。3 个 α-螺旋伸展进入大沟的一个转折处。每个 α-螺旋（即每个锌指）与 DNA 形成两个序列特异性的接触（图中箭头所指处）。每个锌指 C 端侧的非保守性氨基酸可能负责识别特异的目标位点。

已知的锌指结构主要发现于促进 RNA 聚合酶 II 和 RNA 聚合酶 III 转录的转录因子中。当一个蛋白质有多锌指结构，至少具有了为进一步研究确认其可能具有转录因子作用的初步证据。这样的研究发现了黑腹果蝇的胚胎发育有关的几个基因座表达产物是转录调节物。但是这种根据存在锌指结构进行的推断应该慎重，特别是当蛋白质含有仅一个锌指结构时更

应慎重。锌指结构也可能是结合 RNA 而不是 DNA，也可能与核酸结合活性无关。

典型的锌指蛋白——TFⅢA 与 5 S rRNA 基因及其产物 5 S rRNA 都结合。翻译起始因子 eIF2β 上有一个锌指结构，其锌指结构突变影响对起始密码子的识别。逆转录病毒衣壳蛋白有一个与锌指有关的基序，可能参与同病毒 RNA 的结合。

类固醇受体及一些其他蛋白质有另一类锌指。其结构是以下列锌指结合共有序列为基础。

$$Cys\text{-}X_2\text{-}Cys\text{-}X_{13}\text{-}Cys\text{-}X_2\text{-}Cys$$

此锌指结构被称为 Cys_2/Cys_2 锌指，以区别于 Cys_2/His_2 锌指。与含串联重复 Cys_2/His_2 型锌指的蛋白质不同，含 Cys_2/Cys_2 锌指的蛋白质中的锌指常常不是重复的。其 DNA 结合位点较短，且呈回文结构。调节蛋白突变分析表明它们结合 DNA 的位点包括锌指基序。

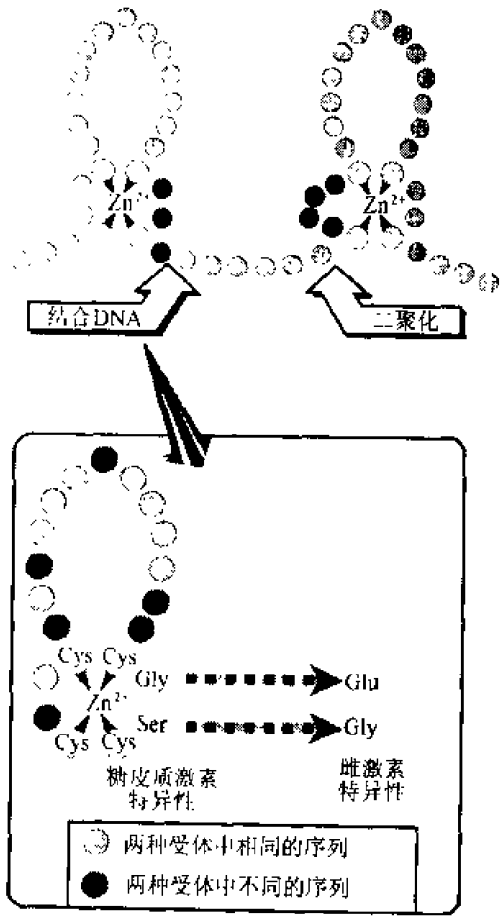


图 8－4　类固醇受体两个锌指的功能

上图示意糖皮质激素受体第一个锌指控制 DNA 结合的特异性，第二个锌指控制形成二聚体的专一性。下图显示与糖皮质激素受体和雌激素受体所结合的靶序列 GRE 和 ERE 对应，它们第一个锌指上氨基酸的差异。

糖皮质激素受体和雌二醇受体都有两个锌指结构，每个在半胱氨酸四面体中央都有一个锌原子。两个锌指结构形成 α－螺旋，再折叠形成一个大球形结构域。α－螺旋的芳香侧链与连接两个螺旋的 β－折叠一起形成疏水中心。螺旋的 N 末端侧与 DNA 的大沟接触。

两个糖皮质激素受体借与 DNA 的结合形成二聚体，每个都占据大沟一个连续转折。这与效应元件的回文结构的性质相吻合。

每个锌指控制受体的一个重要性质。图 8－4 说明了与此相关的氨基酸。第一个锌指的右侧控制 DNA 的结合，第二个锌指的左侧控制形成二聚体的能力。第一个锌指结合 DNA 的直接证据已从“特异性交换(specificity swap)”实验得到。将雌激素受体的锌指缺失并用类固醇受体代替，新的蛋白质识别糖皮质受体的靶序列——GRE，而不识别雌激素的靶序列 ERE。说明此锌指决定 DNA 识别的特异性。糖皮质激素受体与雌激素受体的主要差异位于锌指区，图 8－4 中所示的两个氨基酸的改变就会改变两者的结合 DNA 上调控元件的特异性。

五、类固醇受体的几个独立结构域

类固醇激素在各种神经内分泌活动控制下合成，在动物的生长、组织发育和身体稳态等方面发挥重要作用。类固醇及有相关活性的激素

列于图 8－5。

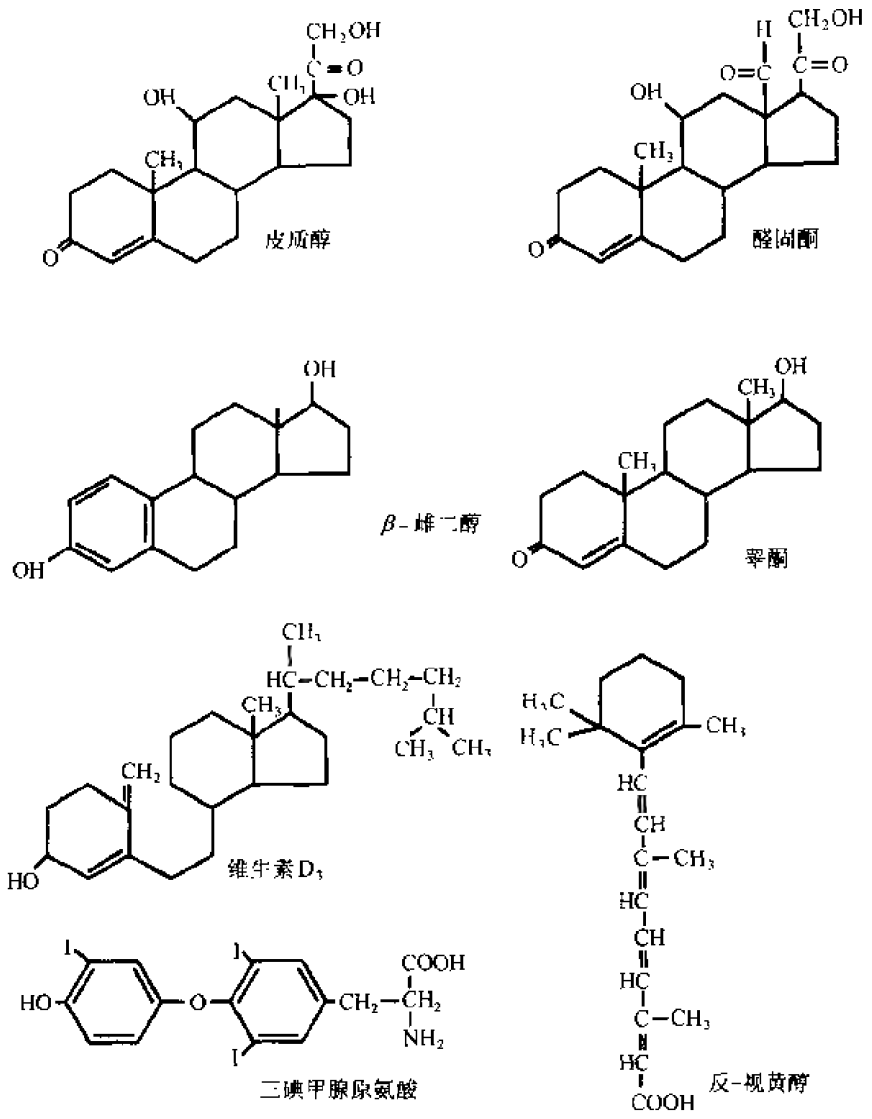


图 8－5　几种能激活转录因子的疏水小分子

肾上腺皮质激素和性激素由胆固醇合成，维生素 D 是类固醇，甲状腺激素由酪氨酸合成，视黄酸由异戊二烯（在鱼肝中）合成。

肾上腺分泌 30 多种类固醇，其中最主要的两组是糖皮质激素和盐皮质激素。类固醇也包括生殖激素——雄激素（androgen）和雌激素（estrogen）。骨骼发育则需要另一种类固醇——维生素 D。另一种激素结构和生理功能与类固醇激素不相同，但是在分子水平上，以相似的方式发挥作用，它们是甲状腺素。甲状腺素酪氨酸的碘化物，控制着大鼠等动物的基础代谢。类固醇和甲状腺素在一些动物变态中有重要作用。如蜕皮类固醇激素和甲状腺素分别在昆虫和蛙类的变态中发挥作用。

视黄酸（维生素 A）是一种器官发生素，负责小鸡肢芽体发育中的前后轴的发育。维生素 A 的代谢物 9－顺视黄酸主要存在于贮存和代谢维生素 A 的组织中。

我们可以把调节身体发育和功能的各种作用称为基因表达调节途径。这些化合物有着相同的作用模式：即都是与特异受体结合的小分子，它与受体的结合激活基因转录。此

处用“受体”这个词似乎用词不当。作为类固醇和甲状腺素受体的这些蛋白质与作为β-半乳糖苷“受体”的乳糖操纵子阻抑物含义相同。它不是含义为暴露于细胞表面的膜结合蛋白的“标准”受体。

如图 8-6 所示，我们对糖皮质激素与其受体的相互作用有了相当程度的了解。类固醇激素能经简单扩散穿过细胞膜进入细胞核，在细胞内与其受体结合。游离受体的细胞内定位仍不完全清楚，可能在细胞核和细胞浆内维持平衡。但是当激素与受体结合时，受体蛋白质便转变成活性形式，对非特异 DNA 结合的亲和力也提高 10 倍，因此激素-受体复合物总是位于细胞核内。

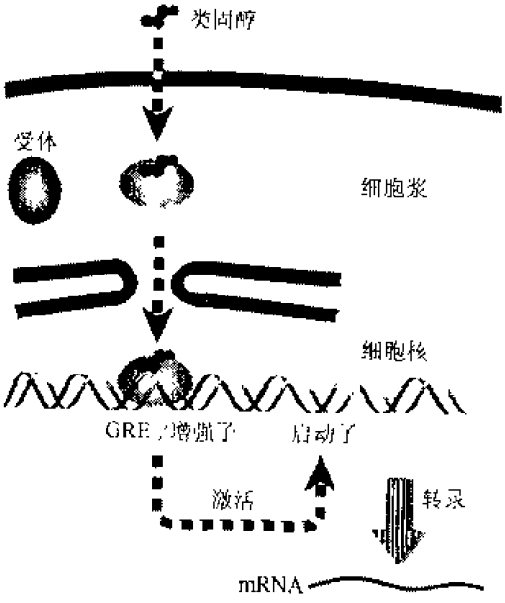


图 8-6 糖皮质激素对转录的调节
糖皮质激素进入细胞内，与受体结合，激素-受体复合物进入细胞核，与增强子结合，从而激活启动子，开始转录。

激活的受体识别其特异共有序列，该共有序列就是糖皮质激素效应元件（GRE）。GRE 总是位于应答糖皮质激素的基因附近的增强子内。GRE 可能位于启动子上游数千碱基对处，也可能位于启动子下游几千碱基对处。当类固醇受体复合物与增强子结合时，激活附近的启动子，使启动子开始转录。类固醇激素调节一系列靶基因都是通过激活增强子实现的。

各种类固醇激素、甲状腺素和视黄酸形成一新的基因转录调节物“超家族（superfamily）”。它们都属配体应答性的转录因子。所有的受体都有独立的 DNA 结合和激素结合结构域。各种受体上这些结构域的相对定位相同。

各种类固醇受体 DNA 结合域的中心结构很相近，且与其他两类受体（甲状腺素和视黄酸的受体）相似。一些配体有多个受体，这些受体间的 DNA 结合域很接近。如 3 种视黄酸受体（RARα、β、γ）和 9-顺视黄酸的 3 种受体（RXRα、β、γ）间的 DNA 结合域就特别接近。这些结构域上序列的相同点反映出其结合的 DNA 的共同点，其差异则使他们能结合各自的靶序列。DNA 的结合作用与其激活转录的能力有很密切的关系。对 DNA 结合域的突变会影响受体的结合活性及激活活性。

各种受体的 N 末端区同源性最低。该区含转录所需的另一些区域。

受体的 C 末端区与激素结合，类固醇受体家族各受体的此区域与糖皮质激素的同源性为 30%～57%。类固醇受体的此区与甲状腺素、维生素 D 和视黄酸等的受体同源性小，表现出激素的特异性。

由于不同激素受体 C 末端区不同，从而决定了不同激素有不同的活性。如果将糖皮质激素受体 C 末端缺失，保留的蛋白质具有组成型活性，即不再需要类固醇激素就有活性。这表明，类固醇激素的缺乏，其受体的类固醇结合域阻止受体识别 GRE，从而起内在负调控物的作用。加入类固醇能消除其抑制作用，使受体能与 GRE 结合，激活转录。

而雌激素受体的结构域间的关系与此不同。如果缺失雌激素受体的激素结合域，尽

管仍能与 ERE 结合，但受体蛋白质不能激活转录。因此雌激素受体的激素结合区具有激活活性，而不是阻抑活性。

每种受体识别的效应元件很保守。这些效应元件有很明显的特点：每个共有序列都含两个短的重复序列，每个重复序列也可称为半位点。这表明受体有多个亚基，每个半位点与一个亚基接触。

各种受体的效应元件也可能是反向或正向重复序列。每个重复序列中，半位点被 0～4 bp 分隔，分隔的这些碱基序列则没有特别的要求。各种受体所用的半位点仅有两种类型。其方向和间隔决定了是哪种受体的识别元件。这使得各种受体特异识别的效应元件是限制性的共有序列。这种规则决定了识别不是绝对的，而是可以在上下文中进行调整，也有反向重复的效应元件可被多个受体识别的情况。

受体可分为两类：

第一类是糖皮质激素受体(GR)、盐皮质激素受体(MR)、雄激素受体 AR)和孕酮受体(PR)，它们都是相同亚基组成的二聚体，所识别的元件的半位点有共有序列 TGTTCT。半位点反向重复排列着，位点间的间隔决定着其识别元件的类型。雌激素受体(ER)也以同样的方式发挥作用，但是其半位点序列是 TGACCT。

第二类是甲状腺素受体(T3R)、维生素 D 受体(VDR)、视黄酸受体(RAR)及 9-顺-视黄酸受体 RXR)。这类受体除 RXR 外都由两个不同的亚基构成二聚体，所识别的元件的半位点序列是 TGACCT。半位点直接重复排列，识别特异性由半位点间的间隔序列长短决定。间隔序列长度与对应的受体关系是：1 bp 对应 RXR，2 bp 对应 VDR，3 bp 对应 T3R，4 bp 对应 RAR。

到此，我们可以理解识别特异性的基础。我们再回头看看图 8-4，由第一指的氨基酸序列推测其怎样识别半位点序列。二聚体形成的特异性由第二指的氨基酸序列决定。二聚体一形成就确定了亚基间的距离，亚基位于大沟的连续的转折内，因此控制着对半位点间隔的应答。

由相同亚基(GR 或 ER)或不同亚基(VDR、T3R 或 RAR)构成的二聚体的结构当然有差异。由相同亚基构成的二聚体形成“头对头”的对称结构，识别 DNA 的反向重复序列。由不同亚基构成的二聚体形成“头对尾”的不对称结构，其中二个亚基聚合时的界面决定了识别的 DNA 重复序列的间隔。

所有的超家族的受体都是配体依赖性的转录激活物。然而，有些也能阻抑转录。TR 和 RAR 与 RXR 形成异源性二聚体，在配体存在时结合一些基因座，借其与辅阻抑物蛋白相互作用的能力阻抑转录。

六、结合相关靶 DNA 的同源域

同源框(homeobox)是一种编码由 60 个氨基酸组成的结构域的序列，它存在于许多或几乎所有的真核生物的 DNA 结合蛋白中。由于最早在果蝇的同源框基因座(其基因决定身体结构的特点)中发现而得名。它存在于果蝇的很多调节早期发育的基因中，在各种更高等的真核生物中的基因中也发现了类似的基序。同源框与发育调控有关。在一些类型的动物转录因子中也发现了与同源框类似的序列，但是到哺乳动物转录因子，此序列

与果蝇的同源框有了很大差异。果蝇的同源框基因中，其编码的蛋白质的同源（异型）域常紧靠着 C 末端。各种同源框基因除了同源框本身外，其余部分保守性很低。同源框共有序列有多种。果蝇中一大组含同源框的基因的同源框有较好的保守性，同源性达 80%～90%。其他基因有同源性较差的同源框。在动物的转录因子中，同源（异型）域有时与其他基序结合。这种情况见于 Oct 蛋白（八核苷酸结合蛋白）中，该蛋白中存在一含 75 个氨基酸残基的 Pou 区与同源（异型）域相邻。

同源（异型）域负责与 DNA 结合。同源（异型）域的 C 端区与原核生物阻抑蛋白的螺旋-转折-螺旋基序同源。λ噬菌体阻抑蛋白有一个“识别螺旋”（α-螺旋-3）与 DNA 大沟接触，而另一个螺旋（α-螺旋-2）位于单过 DNA 的角。同源（异型）域可能有三个螺旋区，三种蛋白的这些区域见图 8-7。其中保守性最好的位于第三个螺旋中。这些结构与原核生物阻抑蛋白结构的差异在于识别 DNA 的螺旋长度不同。如螺旋 3，在同源（异型）域中为 17 个氨基酸残基，而在λ噬菌体的阻抑蛋白中为 9 个氨基酸残基。

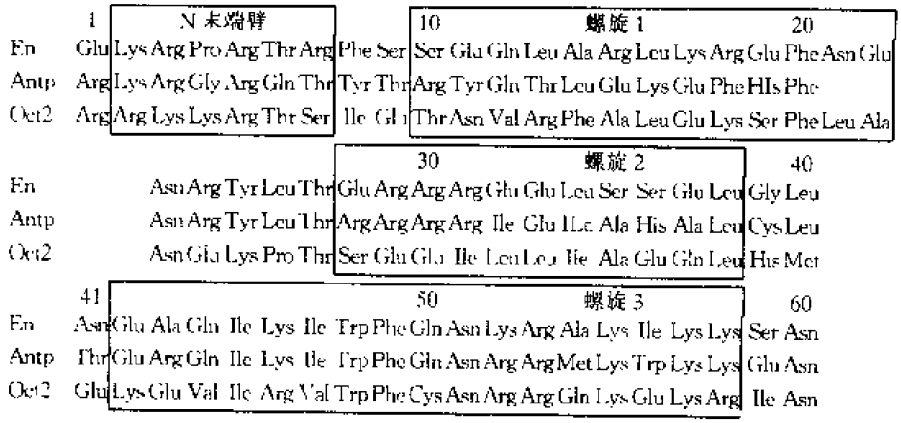


图 8-7　一些蛋白同源异型域的三个螺旋区

Antp 是果蝇的触角足（antennapedia，Antp）蛋白的同源（异型）域，代表果蝇中含同源（异型）框的大多数蛋白质；En 是果蝇的 engrailed，代表另一型的含同源（异型）框的蛋白；Oct2 是哺乳动物因子 Oct2，代表一组远缘相关的转录因子。图中列出的同源（异型）域的 60 个氨基酸，编号为 1～60。开始于 N 端臂，三个螺旋区分别为 10～22、28～38 和 42～58。

七、螺旋-环-螺旋结构

DNA 结合蛋白的两个共同特点是存在与 DNA 结合的螺旋区及能形成二聚体。在一类螺旋-环-螺旋的蛋白中两个特点都有，这些蛋白质有一共同类型的基序：此基序长 40－50 个氨基酸残基，其中含两个既亲水又亲脂的 α-螺旋，α-螺旋被不同长度的连接区（环）分开。此类蛋白质借两个螺旋对应面上疏水基团相互作用形成同样亚基或不同亚基构成的二聚体。螺旋区长 15～16 个氨基酸残基，每个都有几个共有的氨基酸残基。图 8-8 对两个例子进行了比较。这些亲水亲脂螺旋的残基形成二聚体的能力与 HLH 蛋白质相同。环也许对自由的两个螺旋区相互作用很重要。

大多数 HLH 蛋白有一与 HLH 基序相邻的强碱性的区域，它是与 DNA 结合必需的。此区的 16 个氨基酸中有约 6 个共有氨基酸残基。含有此区的 HLH 称为 bHLH 蛋白。由同样亚基组成的二聚体及由不同亚基组成的二聚体的每个亚基都有与 DNA 结合的碱

性区。HLH区可能有使各亚基的两个碱性区正确定向的作用。但仍不清楚HLH蛋白怎样激活转录。

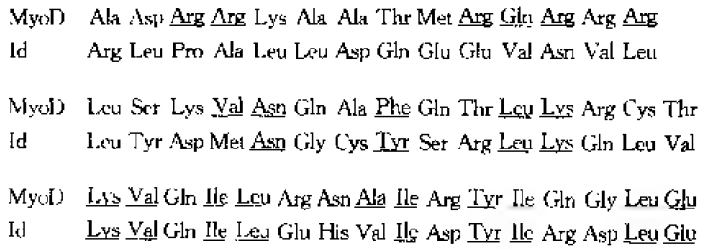


图8-8 MyoD及Id的HLH结构

下划线的氨基酸残基为共有(保守)氨基酸残基。所有HLH蛋白都有由10~24个氨基酸残基构成的环隔开的螺旋1和螺旋2区。碱性HLH蛋白紧靠着螺旋1含保守的正电荷区。

属bHLH类的蛋白有E12和E47、MyoD、肌生成素(myogenin)、Myf-5转录因子、Myc蛋白及黑腹果蝇的神经系统发育特异的基因的产物。E12和E47能与免疫球蛋白基因的增强子的一个元件结合。MyoD、肌生成素和Myf-5转录因子参与肌肉形成。Myc蛋白是癌基因的产物，参与生长调控。

bHLH蛋白可分为两类。A类蛋白呈广泛性表达，其中有哺乳动物的E12和E47，蝇的da；B类蛋白呈组织特异性表达，其中有哺乳动物的MyoD和蝇的AC-S_c。组织特异性的bHLH蛋白与广泛性的bHLH蛋白形成异源体能最有效地发挥转录调控作用。我们不知道是否这是唯一的活性形式，是否广泛性的bHLH蛋白有时也参与调控转录。Myc蛋白是另一类bHLH蛋白，其形成异源体的分子伴侣及作用的对象都不同。

bHLH蛋白形成的二聚体结合DNA的能力不同。例如，E47纯合二聚体、E12-E47异源二聚体和MyoD异源二聚体都很易形成，且都有较高DNA结合活性。E12纯合二聚体也易形成，但DNA结合活性很弱。因此，二聚体的形成和DNA结合可能都是重要的调节点。在此结合点上，有可能对HLH蛋白的各成员形成各种配对结合进行分类，但是从序列上不能推算二聚体形成和DNA结合的程度。所有的二聚体结合的DNA都有同样的共有序列，但是我们并不了解是否纯合二聚体和异源二聚体喜好的目标序列是否明显不同。

DNA结合的差异是由该区内或紧连着的HLH基序的性质决定的。例如，E12与E47由碱性区形成一个抑制区上不同，该抑制区阻止DNA与纯合二聚体结合。一些HLH蛋白缺乏缺乏碱性区，且/或含脯氨酸残基，而脯氨酸残基似乎能破坏其功能。蛋白质Id的例子在图8-8中进行了说明。这类蛋白质有像bHLH蛋白一样形成二聚体的能力，但是二聚体只要含该类蛋白的一个亚基就不能与DNA特异结合。这是DNA结合蛋白质中两份DNA结合基序的重要性的强有力的证明。

从两组HLH蛋白(da-AC-S/emc组和MyoD/ID型的性质可以表明区分非碱性HLH和bHLH蛋白的重要性。它们在形成调控网中的作用见示意图8-9。

黑腹果蝇的*emc* (*extramacrochaetae*，外长毫)基因是建立成年感觉器官正常的空间型所必需的。它的功能是抑制几个基因的功能，受抑制的基因有*da* (*daughterless*)和*achaete-acute*复合体(*AC-S*)，如果这些基因不被抑制，则其他的细胞将得到表达。*AC-S*

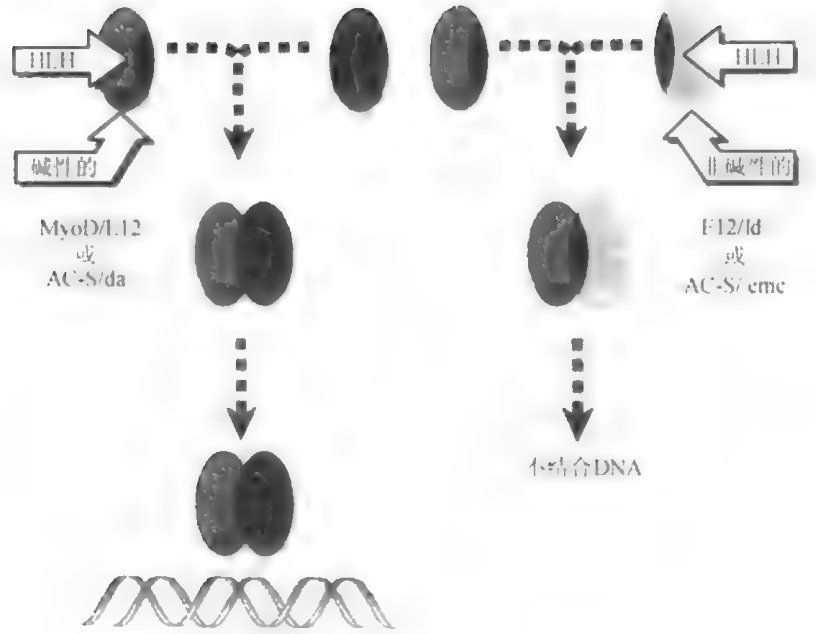


图 8-9 两类 HLH 二聚体的作用

左图示意二个 bHLH 类的亚基形成的二聚体能与 DNA 结合，右图示意只要有一个是非 bHLH 类的亚基形成的二聚体就能与 DNA 结合。

和 *da* 是 bHLH 类基因。抑制基因 *emc* 编码缺乏碱性区的 HLH 蛋白，而 da 和 AC-S 蛋白形成二聚体则能激活适宜的目的基因转录，但是 emc 蛋白参与形成的异源二聚体不能与 DNA 结合，不具有激活转录的作用。因此，在某些特定的细胞内，产生 emc 蛋白以抑制 AC-S/da 的功能是必要的。

一种需要包括 MyoD 等几种 bHLH 蛋白的转录程序变化能引发肌肉细胞的形成。MyoD 只在生肌细胞(myogenic cell)中产生。确实，MyoD 在一些其他细胞中过度表达能诱导这些细胞开始生肌程序。肌肉分化的引发物也许是 MyoD-E12 或 MyoD-E47 组成的异源二聚体，而不是 MyoD 组成的纯合二聚体。在肌生成开始前，非碱性 HLH 型蛋白，Id 蛋白能与 MyoD 结合，并与(或与)E12 和 E47 结合，形成不能与 DNA 结合的异源二聚体。Id 蛋白结合 E12/E47 较结合 MyoD 紧，表明 Id 可能是起阻隔广泛存在的 bHLH 分子的作用。Id 的过度表达抑制肌生成作用。

因此，HLH 蛋白的行为说明了转录调控的两个一般原理。即少量的蛋白质形成复合物；有可能特定的结合对 DNA 结合和转录调节有不同的功能。但是我们仍不能完全理解组织特异性的及广泛存在的 bHLH 的功能。此外，大量相同的通用类的蛋白质通过参与形成复合物而发挥阻抑物的作用。特定的非碱性 HLH 类阻抑蛋白的存在和去除决定了分化。

八、亮氨酸拉链结构

蛋白质相互作用是形成转录复合物的基本规则，转录因子中的结构域参与蛋白质间的相互作用。亮氨酸拉链是一个富含亮氨酸残基的结构域，参与形成二聚体。二聚体形

成本身作为蛋白质识别特异 DNA 序列的一般规则出现，在亮氨酸拉链结构中，这与 DNA 结合的关系更是清楚，因为我们能看见 DNA 怎样与二聚体中的每个亚基进行结合。图 8－10说明了二聚体中各亚基中的亮氨酸拉链与 DNA 的这种结合作用。

两亲性的 α－螺旋具有这样的结构：疏水基团（包括亮氨酸）面向一侧，而带电荷的基团面向另一侧。亮氨酸拉链形成两亲螺旋，其中，一个蛋白质拉链的亮氨酸能突出于 α－螺旋，并与另一个平行的蛋白质拉链上的亮氨酸交错对插，盘绕成卷。两个右手螺旋互相缠绕，每圈含 3.5 个氨基酸残基，因此，每 7 个残基构成一个完整的重复单位。

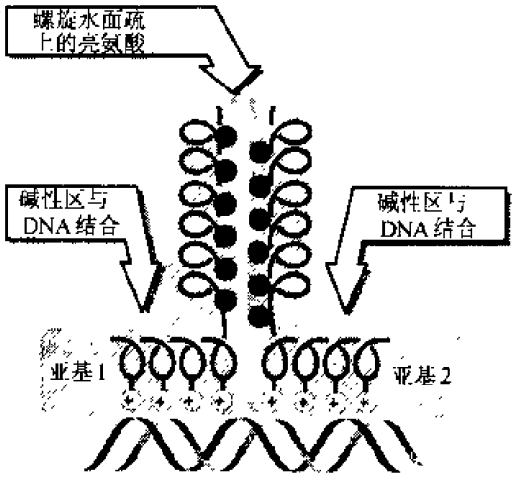


图 8－10　亮氨酸拉链参与亚基二聚体形成示意图
当两个亮氨酸拉链的疏水面平行排列时，碱性亮氨酸拉链基序的碱性区维持相邻的拉链区的聚合。

在每个拉链蛋白中与重复的亮氨酸相邻的高碱性区，该区可能含一个 DNA 结合位点。两个亮氨酸拉链作用形成一个 Y 型结构，其中，拉链构成茎，两个碱性区分叉对称形成臂，与 DNA 结合。这种结构就称为 bZIP 基序。它解释了为什么这种蛋白质的目标序列总是没有间隔的反向重复序列。

拉链可能协助形成纯合或异源二聚体。它们是冗长的基序。亮氨酸在拉链每隔 6 个氨基酸残基出现一次。在 C/EBP 上有 4 个重复单位，在 Jun 和 Fos 因子上有 5 个重复单位。C/EBP 是与 CAAT 框及 SV40 核心增强子都能结合形成二聚体的因子。Jun 和 Fos 因子与转录因子 AP1 能形成异源二聚体。

AP1 最初是因其能与 SV40 增强子中的 DNA 序列结合而鉴定得到。AP1 的活性制品包括几个多肽。一个主要成分是 Jun，即基因 c-*jun* 的产物。c-*jun* 是根据其与禽肉瘤病毒所含的原癌基因 v-*jun* 的关系而确定出的。小鼠基因组含一个相关基因的家族，它们是 c-*jun*（最早分离到的）及 *junB* 和 *junD*（根据其与 *jun* 序列的同源性鉴定出的）。在三种 Jun 蛋白质中，有相当多的序列相似，他们有能形成纯合二聚体及异源二聚体的亮氨酸拉链。

AP1 的另一个主要成分是原癌基因对应的另一个基因的产物。这个基因是 c-*fos* 基因，他是小鼠肉瘤病毒原癌基因 v-*fos* 的细胞内的同源物。c-*fos* 的表达激活含有 AP1 位点的启动子或增强子的基因的表。c-*fos* 基因的表达产物是一个核磷蛋白，这种蛋白是一类蛋白的其中一个蛋白；另一个被称为 Fos 相关抗原（Fos-related antigens，FRA）；这些蛋白质组成 Fos 样蛋白家族。

Fos 蛋白也有一个亮氨酸拉链。但是 Fos 不能形成纯合二聚体，而只能与 Jun 形成异源二聚体。为了反应，每个蛋白质都需要有一个亮氨酸拉链。形成二聚体的能力是这些因子与 DNA 相互作用的关键因素。Fos 蛋白本身不能与 DNA 结合，这可能是因为它不能形成二聚体。但是 Jun-Fos 异源二聚体能与 Jun-Jun 二聚体结合同样的靶序列；且异源二聚体结合 AP1 位点的亲和力是 Jun-Jun 二聚体的约 10 倍。

九、基因激活占先模型的动力学

到此为止，我们已经讨论了DNA和转录因子及RNA聚合的相互作用。这在体外实验中得到了精确描述，但缺乏体内转录的证实。细胞内基因组包裹在核小体内；结果启动子区在核小体内，转录的起始通常会被抑制。此时，组蛋白起一般转录阻抑蛋白的作用，尽管我们在下面看到也有更多的特异性作用。一个基因的激活需要改变染色体的状态；即转录因子怎样得到启动子DNA并与之结合。

这与细菌细胞的转录调控相差较大，在细菌细胞中，主要通过阻抑蛋白或激活蛋白发挥调节作用。一个细菌的基因转录能根据各种因子的浓度进行预测，或是阻抑，或是激活。任何时间，这些因子浓度的改变都会改变表达调控状态。

这种简单的调控关系不能用于真核生物的染色质状态对调控的影响。可以想象，大量的特异性转录因子可以取代靶启动子上的核小体，但简单的平衡模型很难解释在整个细胞周期中基因的稳定表达，因为在细胞周期的S相中，合成了大量组蛋白。

有两种改变染色质结构的模型。一种是占先模型。该模型认为，为了占有DNA，转录因子与组蛋白之间存在竞争，且无论谁先占领了DNA上的位点，都不能被另一方替换。复制时组氨酸八聚体脱离为转录因子结合DNA提供了机会，这些转录因子的结合一直持续到下一个复制周期，抑制了组氨酸八聚体的重新与DNA结合。

早期实验所得结果支持此模型。负责转录5 S rRNA基因的RNA聚合酶III需要转录因子TFIIIA，如果5 S rRNA基因与组蛋白形成复合物，则转录因子不能激活此基因。然而转录因子能与游离DNA形成必要的复合物，且此复合物形成后，再加入组蛋白不能使基因失活。一旦转录因子与其位点结合，就使得RNA聚合酶能起始转录。因此起决定的因素是转录因子和组蛋白两者谁首先与控制位点结合。

与此类似的现象在RNA聚合酶II启动子的TFIID复合物上也可见到。含腺病毒启动子的质粒能在体外被RNA聚合酶II转录，此转录反应需要TFIID及另一些转录因子参与。加入组蛋白能将转录模板装配到核小体上。如果组蛋白先于TFIID加入，则转录不能起始。但如果先加入TFIID，则在形成的染色质中，模板仍能转录。表明，TFIID能识别游离DNA，不能识别核小体上的DNA。仅有TFIID必须在组蛋白前加入，其他转录因子和RNA聚合酶在组蛋白后加入也有活性。这说明TFIID与启动子结合形成了一个其他转录装置成分能结合的结构。

此模型的一个重要特点是要不形成核小体构型，DNA上必须存在转录因子。如果DNA复制时没有某种转录因子，核小体形成，不能转录。在这些结合DNA的占先机制的例子中，转录因子不能与已形成了核小体的DNA有效结合。然而在解释这些体外实验得到的结果时，应该注意到所用的各种成分的量的比例不一定合适，这种比例可能会扰乱应有的平衡，产生新的平衡。所以这些结果的重要之处不在于它说明体内采用怎样的机制，而是说明了转录因子一旦存在，就会使启动子区的核小体不能形成这一原理。

最近的实验结果认为组蛋白置换需要输入能量。大量的蛋白质与蛋白质及蛋白质与DNA接触需要组蛋白从染色质中裂解释放出来，这使得游离组蛋白似乎不可能平衡形成八聚体。然而事实上没有游离的组蛋白，能量一定来自这些连接的破坏。转录因子水解

ATP 可用动力学模型(dynamic model)解释。

果蝇 *Hsp70* 启动子上的核小体在体外实现了重建。转录因子 GAGA 与启动子中 4 个富含$(C\ T)_n$的位点结合破坏了核小体,形成一高敏感区,一旦转录因子从随机位点移到特异位点上,就引起邻近的核小体重排。核小体的这种打开过程是需要水解 ATP 的耗能过程。表明 GAGA 在结合位点上裂解了核小体,边界的形成引起重新排列,使得邻近的核小体变化,成为可结合转录因子的构型。GAGA 甚至在核小体安装后加入也能发挥作用,表明它能改变存在着的核小体结构。但 ATP 缺乏,其功能丧失,表明其发挥作用的需能动力学机制。

十、基因表达与去甲基化的关系

DNA 的甲基化发生在特定位点上。在细菌中,它与鉴定特定菌株有关。也可区分复制和非复制 DNA。在真核生物中,其主要功能是参与转录控制。动物细胞 DNA 的胞嘧啶 2%～7%发生甲基化。大多数甲基化基团发生于 CG 联体,事实上,大多数 CG 序列都被甲基化。通常,在此短的回文结构中 C 残基甲基化,形成如下结构:

$$5'^{m}CpG\ 3'$$
$$3'\ GpC^{m}5'$$

一个 CG 联体中,两条链中只有一条链发生了甲基化称为半甲基化。甲基化基团的分布要用靶序列含 CG 联体的限制性内切核酸酶酶解进行测定。限制性内切核酸酶 *Hpa* II 的识别靶序列为 CCGG。但是如果第二个 C 发生了甲基化,则不再识别此序列。然而其同裂酶 *Msp* I 切割相同的靶序列,对无论是否甲基化的底物都能切割。对于一个非甲基化的底物 DNA,两个酶形成相同的限制性片段,而对甲基化 DNA,*Hpa* II 对甲基化位点不能切开,对每个这样的位点,*Hpa* II 酶解仍得到一个较大的片段,而 *Msp* I 酶解则得到两个片段。

许多基因的大多数甲基化位点的甲基化状态不变,但有少数甲基化位点的甲基化状态会改变。一些位点在所有组织中都发生甲基化,一些位点在所有组织中都未甲基化。少数位点在基因不表达的组织中甲基化,但在表达的组织中不甲基化。因此活性基因称为甲基化不足(undermethylation)基因。

将甲基化及未甲基化 DNA 引入到新的宿主细胞中,可测定"常居"基因的甲基化状态。这些实验得到了明确结论,即甲基化基因没有激活,而未甲基化基因有表达活性。

十一、甲基化确定印记

甲基化状态是怎样能够恒久持续下去或是怎样发生改变的,已引起了极大的关注,对在所有组织中被表达的基因来讲,可能没有什么问题,它们会以甲基化的状态恒久的存在。但受组织特异性控制支配的基因甲基化状态会表现出改变。它们在精子中常常呈现灭活状态;在恒定位点(在所有组织中都被甲基化)和可变位点(在表达的组织中特异性的脱甲基化)都被甲基化。这样,在活性状态,某些甲基的缺乏则意味着失去以前出现过的修饰。

我们不知道当细胞内的基因被停止表达时,是否它们能重新得到甲基。我们要回答

的一个主要问题是，在甲基化状态，序列是怎样被选作组织特异性改变的靶子。

对甲基化位点能够恒久存在的简单模型是假设 DNA 甲基化酶作用于半甲基化的 DNA。全甲基化的 CG 对的复制能够产生两个半甲基化的子双链 DNA，酶能够识别每个半甲基化的位点，将其转化为正常的全甲基化状态。

这个模型与以下观察一致，当将甲基化的 DNA 引入细胞内时，它将通过无限的复制周期来继续被甲基化，这种复制的保真度为每个位点约 95%。如果将非甲基化的 DNA 引入细胞内，则非甲基化的 DNA 不被全甲基化。这意味着酶只识别半甲基化位点。它的作用使$\begin{matrix}CG\\GC\end{matrix}$联体——不管甲基化还是非甲基化——恒久存在下去。

如果这个模型正确的话，在建立甲基化的新位点中（例如，如果当转录停止时一个基因的甲基化状态被重建）必定还有另外一种酶活性参与。以下两种方法中的任意一种可以失去甲基化状态，用脱甲基酶除去活性的甲基，或在通过复制产生的一个半甲基化位点上，甲基化即会简单的停止。下次复制产生任一 DNA 链即会缺失甲基化位点。

使用药物 5-氮杂胞苷进行的实验为甲基化能够导致基因表达提供了间接的证据。药物结合到 DNA 中胞苷的位置而不能被甲基化，因为 5′位置被阻断。这样可以产生 DNA 的脱甲基位点。5-氮杂胞苷的表型作用包括在细胞内分化的状态中诱导变化，例如，从非肌肉细胞前体 180℃诱导肌肉细胞发育，药物还可以激活沉默 X 染色体上的基因，这提高了甲基化状态与可能与染色体失活有关的可能性。

甲基的一种特定类型可以由细菌细胞遗传，这可以解释印记（imprinting）现象，印记描述由每个亲本遗传而来的等位基因之间的行为差异。在鼠胚胎中某些基因的表达取决于其亲本的性别。例如，源于父本的编码 IGF-II（胰岛素样生长因子 II）的等位基因能够被表达，而源于母本的则不能被表达。最可能的解释是，卵母细胞的 IGF-II 基因已甲基化，而精子的 IGF-II 基因未被甲基化，所以这两种等位基因在接合子中的表现不同。这是最常见的形式，对某些基因，对性别的依赖会出现反转。

这种模型要求甲基化的类型要建立在特定的配子发生期间，而且必须是可以反转的。在早期胚胎，父本等位基因未被甲基化而被表达，母本等位基因被甲基化而沉默。当这只鼠自身形成配子会出现什么情况？如果这只鼠是雄性的，在形成的精子中的等位基因一定未被甲基化，而与它开始时是否被甲基化无关。这样，当母源等位基因进入精子时肯定被脱甲基化。如果这只鼠是雌性的，在形成的卵子中的等位基因一定被甲基化，所以，如果等位基因开始时是父本的，则必定要加上甲基。这样，我们要求一个系统能通过重新除去甲基或加上甲基，特异性地重置甲基化状态。

甲基化模式的改变发生在胚胎发生时期。当最初的精子细胞在胚胎中发育时，所有等位基因的差异都消失了，不管是什么性别，以前的甲基化模式被抹去，然后一种典型的基因被脱甲基化。

在每种性别的配子发生的不同时期都确立了性别的甲基化模式的差异。在雄性，这种模式在两个时期中发育。精母细胞表现出成熟精子特点的甲基化模式，所以在精子发生时不出现进一步的变化。但在受精后这种模式将出现进一步的变化。在雌性，当分娩后卵母细胞通过减数分裂而成熟时，在卵子发生期间，母源模式一直保持。

只确定了有一种单一的甲基酶将甲基添加到 CpG 对上。它是一个必需基因，因为这

种基因受到破坏的鼠胚胎存活过不了早期胚胎发生期。我们可以推测，适当的靶基因的甲基化对防止它们的不适当表达是必需的。没有有关负责除去甲基的酶的资料。一个主要的问题是在雄性和雌性配子中甲基化的特异性是如何确定的。

第三节　mRNA 前体的可变剪接[1,2]

真核生物编码蛋白质的基因经转录产生的初始转录物需经过加工后，才产生出成熟的 mRNA，然后 mRNA 进入胞浆，经翻译产生出特定的蛋白质。mRNA 前体的这种加工包括 5′加帽、3′加尾、剪接和编辑等。这些加工过程的都受到调控。例如，通过不同的剪接可由一个基因的转录物产生出不同的成熟 mRNA，从而翻译出不同的蛋白质。又如，通过编辑后 mRNA 的密码子有所改变，产生出与基因编码有所不同的蛋白质等。这里先介绍 mRNA 前体的可变剪接。随后再介绍 RNA 编辑。

mRNA 前体中的内含子边界 5′的供体和 3′的受体序列被识别，内含子在剪接过程中应当依次被删除。在线性分子中，剪接体能将 5′供体与其最近的 3′下游的受体结合以保证剪接的正确和完全。然而，高等真核细胞中来自一个基因的 mRNA 前体分子中一般不止一个内含子，因此某个内含子 5′的供体又可以在特定条件下与另一个内含子的 3′受体进行剪接，从而同时缺失这两个内含子及其中间的全部外显子或内含子，这种按不同方式对 mRNA 进行的剪接称为可变剪接(alternative splicing)。来自一个基因的 mRNA 前体因可变剪接而产生多种 mRNA，翻译出不同的蛋白质，或形成一组相似的蛋白质家族，都可称为同工型蛋白质(isoform)。同工型蛋白质可以在不同的发育阶段、不同的组织中，或在细胞内不同的亚细胞结构中出现并发挥其功能。

一、可变剪接的方式

可变剪接可因 mRNA 前体的外显子或内含子 DNA 序列中发生突变、缺失，影响 5′或 3′剪接点的数目和位置。此外，与 mRNA 前体中内含子剪接位点结合的各种 snRNP 中，snRNA 或蛋白质的异常则会导致剪接效率大大降低。有些基因常不止一个转录起始位点，在不同的组织中两个起始点之间可间隔数千碱基，其中并含有内含子，因此就会导致同一个基因转录出两种 mRNA 前体，需要以不同的剪接方式产生有活性的蛋白质。有的基因的 3′端常有多个加多聚腺苷信号，有可能产生不同 3′外显子的的蛋白质同源体。此外，还有选用不同的剪接位点，产生不同的蛋白质的可变剪接方式。

二、可变剪接的调控机制

1.SR 蛋白家族的调控

剪接体的形成与多种蛋白质和 RNA 复合体密切相关。在可变剪节上保守的 SR 磷蛋白家族因子的参与起重要作用。首先，SR 蛋白早期结合 mRNA 前体决定 5′剪接位点。SR 蛋白强化剪接元件与 5′剪接位点的结合可能基于稳定 U1 snRNA 与 mRNA 前体的结合，或诱导 mRNA 前全的二级或三级结构改变以利剪接前体其他成分的结合等作用。在

可变剪接中 SR 蛋白强化 5′剪接位点的"近端效应"起主导作用。通常 5′剪接位点的突变会激活附近的隐蔽剪接位点，然而 SR 蛋白不能激活隐蔽的 5′剪接点。SR 蛋白家族以不同质的组成与量的差异选择特定的 mRNA 前体剪接位点，不同的 SR 家族蛋白对适当的 mRNA 前体可以诱导出不同选择的剪接方式，在选择剪接上起决定性作用。SR 家族中任何成员可以恢复缺乏 SR 蛋白的体外剪接活性　SR 蛋白在远距离激活剪接时很可能通过蛋白质之间的相互作用，该蛋白的 R-S 结构可能参与这一机制。

2. RNP 的调节

(1)hnRNP-A1 的作用　　hnRNP-A1 在不同组织中的含量变化很大，不参与组成性剪接。它与 SR 蛋白相反，有利于远端 5′剪接点的选择性活化。hnRNP-A1 不抑制 SR 蛋白在组成性剪接中的作用；但在体外可对抗 SR 蛋白家族中 ASF/SF2 和 SC35 等，在选择 5′和 3′剪接点时所具有的可变剪接活性，并且因此成为参与可变剪接的调节因子，并在体内与 SR 蛋白共同调节组织特异性的剪接。

(2)U5－hnRNP 的作用　　在酵母中 U5　snRNP 参与 5′剪接点突变的可变剪接。U5－snRNP 保的 I 环序列是 GCCUUUUAC，其中第 5 位或第 6 位的 U 可用于活化原有切点附近的隐蔽 5′剪接点，如果 mRNA 前体内含子的第一个 G 突变为 A 时，I 环序列的第 3 位 C 和第 4 位 U 还可与 3′剪接位点的外显子中前两个碱基配对，然而这种碱基间的作用都很弱。哺乳动物细胞中 5′剪接点的保守性很低，U5　snRNP 这种很弱的碱基相互影响可能起协助 U1 结合的作用。

3. 外显子限定模型

哺乳动物细胞 mRNA 前体中由于内含子远远大于外显子，很难设想 5′与 3′剪接位点跨越数万个核苷酸准确的组合在剪接体中。然而确实有证据表明在前速激肽原 mRNA 前体的可变剪接中，外显子下游的 5′剪接位点可影响其上游内含子 3′的剪接。

三、反式剪接

前面讨论的剪接过程都是发生在一个 RNA 分子内部，即通过剪接将一个 RNA 分子内的内含子剪掉，使外显子连接在一起，这种剪接方式称为顺式剪接(*cis*-splicing)。已发现，在锥虫、线虫以及植物的叶绿体中发生两个 RNA 分子之间的剪接，即反式剪接(*trans*-splicing)。反式剪接是以两种不同来源的 RNA 前体分子作为底物，经过剪接在成熟的 mRNA 非翻译部分 5′端剪接上一小段称为"剪接前导序列(spliced leader, SL)"或"小外显子(mini-exon)"的 RNA 片段。这些片段本来并不存在于相应的编码基因前面，而是由其他 DNA 链转录而来。已发现的各种 SL RNA 有一些共同的特点，他们都折叠成一种相同的二级结构，此结构有 3 个茎-环(stem-loop)和一个单链区，是以 snRNA 的形式存在的。在真核生物中虽然反式剪接比顺式剪接少得多，但它确实存在。

在锥虫(*Trypanosome*)中发现有许多 mRNA 的 5′端有个共同的 35 碱基的前导序列(leader sequence)。然而此前导序列并不由每个转录单位的上游所编码，而是在由位于基因组其他区域的一个重复单位转录出来的独立的 RNA 中。此 RNA 的 3′端还具有前导

序列以外的一段序列。图 8－11 示意了此 RNA 的组成。紧接 35 碱基前导序列后有一个 5′剪接位点的序列。而在编码 mRNA 的序列中，在成熟 mRNA 的紧接前面有个 3′剪接位点。

如图 8－11 所示，如果前导序列与 mRNA 是由反式剪接连接起来的，则前导 RNA 的 3′区和 mRNA5′区各相当于一个内含子的一半。如果剪接是按通常的核内剪接机制进行的，则首先应在内含子 5′段的 GU 与 3′端 AG 上游的分支点之间发生反应，形成一个 5′－3′连接的磷酸二酯键。此时由于内含子的两个部分不是共价连接在一起的，因而形成字型分子，而不是如通常那样形成套索型分子。已证明此种剪接确实是形成 Y 字型分子，因而支持图 8－11 中所示的剪接模式。

秀丽隐杆线虫（*C. elegans*）中肌动蛋白（actin）基因的表达与上述情况类似，3 种肌动蛋白 mRNA（以及某些其他 RNA）的 5′端是个相同的 22 碱基的前导序列。此序列不被肌动蛋白基因编码，而是由另一个基因编码，此基因独立地转录出一个 100 碱基的 RNA，前导序列是此 RNA 的一部分。

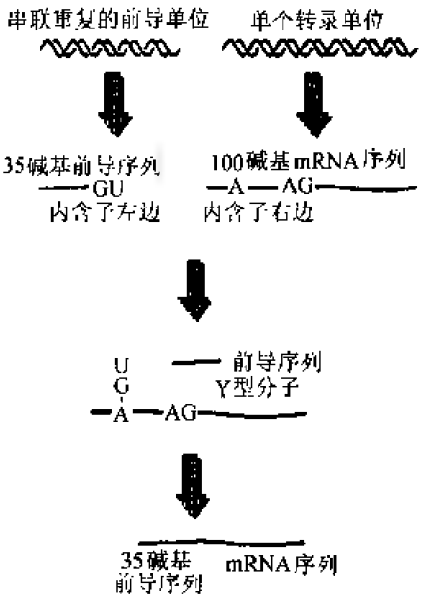


图 8－11　锥虫 mRNA 的反式剪接模式
SL RNA 提供一个外显子，此外显子通过反式剪接与一个 mRNA 的第一个外显子相连接。

某些叶绿体基因是反式剪接的。如单胞藻（*Chlamydomonas*）叶绿体的 *psa* 基因由 3 个相隔很远的外显子所组成。外显子 1 和外显子 2 相距 50 kb，外显子 2 和外显子 3 相距 90 kb。在这些外显子之间有许多其他的基因。而且外显子 1 的转录方向与外显子 2 和 3 的转录方向相反。因此，无论如何他们也不可能是被转录出一个 RNA 分子进行表达的。事实上已鉴定出来的转录只是含有每个外显子（和它两侧的局部序列）的 RNA。因而推断其 mRNA 是由两个反式剪接反应形成的。在这两个反式剪接反应中还需要一些其他的基因，因而可知此 mRNA 的剪接过程是很复杂的。

第四节　RNA 编辑[1,2]

中心法则规定，遗传信息由 DNA 传递到 RNA 过程中，DNA 的序列决定着 RNA 的序列。但是 RNA 编辑似乎是其中的一个例外。RNA 编辑（RNA editing）是在 RNA 分子上出现的一种修饰现象。主要指 mRNA 在转录后因插入、缺失或核苷酸的替换，改变了 DNA 模板来源的遗传信息，从而翻译出氨基酸序列不同的多种蛋白质。编辑的结果不仅扩大了遗传信息，也可能是生物适应的一种保护措施。RNA 编辑的机制如下：

一、核苷酸的替换

1．C→U 替换

最典型的例子是载脂蛋白 B 的 RNA 的编辑。肠型载脂蛋白 B（Apo-B）占肝脏型

Apo-B 100 分子的 48%。在蛋白质水平，它只保留了 Apo-B100 分子 N 端脂蛋白装配结构域，缺少 Apo B 100 分子 N 端脂蛋白装配结构域，缺少 Apo-B100 的 C 端低密度脂蛋白(LDL)受体结合区，Apo-B100 这个 C 端结构域能与 LDL 受体结合，参与胆固醇的运转。在人、兔、鼠中都是由于第 26 个(最大的)外显子中第 6 666～6 668 核苷酸由 Apo-B100 的 CAA 编码的谷氨酰胺突变为 TAA 的终止子，即 C→U 的替换，并产生编码 Apo-B48 的 mRNA。

2. A→I 替换

在爪蟾、自由线虫和某些哺乳动物的培养细胞中，尤其是在血清诱导下，再生长的细胞中都有对同源的 RNA 双链松解和修饰活性的报道。在 β 珠蛋白、c-Myc、成纤维细胞生长因子基因中都有因 A→I，使互补链的 U 被 RNA 酶 A 水解。如果 A 的旁侧为 G 或 C 时，由 A→I 的机率远大于 A 旁为 U 时，后者又稍高于连续为 A 的情况。然而当 A 的上游为 A 或 U 时，A→I 发生的概率也较高。

二、可读框的改变

主要是核苷酸的插入或缺失。插入或缺失 U 只在线粒体、锥虫及其他的动基体中。G 或 C 的插入则往往因为模板上连续几个 C(或 G)之后，互补链的因一种滑动力而被添加到 RNA 中。

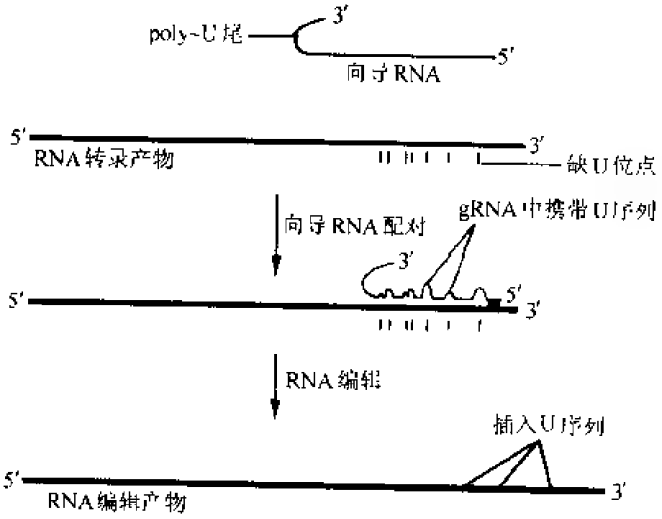


图 8-12 向导 RNA 加 U 机制示意图

三、向导 RNA

向导 RNA(gRNA)是一种线粒体内转录的短 RNA(约 55～70 核苷酸)，能以正常碱基配对或 G-U 配对方式选择其 5′末端“锚”区在 mRNA 上的互补序列，随后为插入或缺失 U 提供模板，此外，在 gRNA 的 3′端有 5～25 个 U，可为 mRNA 增添或接受 mRNA 去除的 U 核苷酸(图 8-12)。

第五节 mRNA 稳定性的调控[1,2]

真核生物 mRNA 的稳定性除了取决于分子内本身的结构特征外，还受转录后修饰(帽子结构、多聚 A 的长短)和与蛋白质所形成的 mRNP 中蛋白质的种类的影响。一些特殊的种类的 mRNA 受特殊因素的影响。如乳腺细胞中酪蛋白 mRNA 的寿命在促乳激素的存在下，稳定性可提高 20 倍。真核生物 mRNA 的比原核生物更加稳定。原核生物

mRNA 的半衰期约为 3 分钟，而在快速生长的脊椎动物细胞，mRNA 的平均半衰期约为 3 小时。

真核细胞中，持家基因的 mRNA 的寿命比较长。因为这些基因的产物是细胞生命活动所必须的。有些蛋白质的 mRNA 分子（如调解细胞周期的蛋白质）的半衰期比较短，因为它们的合成必须以一定的顺序进行。高等真核生物高度分化的细胞中，许多mRNA极其稳定。如在网织红细胞中 RNA 的合成已经停止，但却在进行旺盛的翻译活动，合成的蛋白质 90% 以上都是血红蛋白。由此可见血红蛋白的 mRNA 是非常稳定的。成人肝细胞中参与垂直环系统的蛋白质的 mRNA 也属于这种情况。家蚕丝心蛋白 mRNA 的平均寿命长达几天，一个分子可重复翻译产生约 10 万个蛋白质分子。mRNA 寿命的延长增加了细胞内某种 mRNA 有效浓度，提高了蛋白质合成的速度，这种翻译水平的调控方式称为翻译扩增（translational amplification）。

第六节　翻译的调控

翻译水平的控制是真核生物基因表达多级调控的重要环节之一。翻译的速率和细胞生长的速度之间是密切协调的，当细胞接触到促有丝分裂剂后，蛋白质合成就加快，然而在细胞周期的有丝分裂期时，蛋白质合成就抑制。同样，当细胞受到饥饿、高温或外环境酸碱度的剧烈改变（pH 到达 9 以上或 5 以下），其蛋白质合成也被阻抑。这种改变可在几分钟之内快速出现，然而它是可逆的。参加蛋白质各元件装配活力的改变是导致蛋白质翻译速率变化的主要因素。

一、翻译因子磷酸化调控

蛋白质生物合成起始、延伸及终止的各阶段都有许多因子的参与。仅翻译起始阶段，哺乳动物细胞中就有 13 种因子，酵母也有 12 种因子参与。不少因子磷酸化状态与其对蛋白质合成的激活或抑制作用密切相关。eIF－4F 的磷酸化能激活翻译作用；而磷酸化的 eIF－2α，则能抑制蛋白质合成的抑制。

1. eIF－4F 因子的磷酸化对蛋白质合成的激活作用

在真核生物翻译起始过程中，翻译起始因子 4F（eIF－4F）与 mRNA 5′－m^7G 帽子结构结合，然后 40S－eIF－2－GTP－Met－tRNAi 复合物才能与 mRNA 相连，起始翻译。在哺乳动物细胞中，eIF－4F 由 α、β、γ3 个亚单位，即 eIF－4E、eIF－4A 和 P220 蛋白聚合而成。eIF－4F 通过亚单位的可逆磷酸化对蛋白质生物合成进行调控。当静止期细胞用胰岛素激活后，蛋白质生物合成速度加快。此时 eIF－4F 的 α－亚单位和 γ－亚单位磷酸化作用增加；然而，对细胞进行热激处理，或者细胞处于其生长周期的有丝分烈相时，蛋白质合成受到抑制，同时 eIF－4Fα 亚单位去磷酸化作用。蛋白质合成速率降低与 eIF－4Fα 去磷酸化的相关性证实了 eIF－4Fα 的调控作用。佛波酯 TPA（12　O－tetradecanoyl phorbol－13－acetate）处理细胞增强了 eIF－4F 的磷酸化作用，说明蛋白激酶 C（protein kinase C，PK-C）参与了该因子的磷酸化。

在观察 eIF－4E 对 m^7G 帽子类似物的结合作用时，发现 eIF－4E 的磷酸化和非磷酸化两种形式具有相同的亲和力，这表明该因子的磷酸化不改变对帽子的结合作用。那么 eIF－4F 的磷酸化作用可能是刺激 eIF－4F3 相亚单位复合物形成，或者是促进 含 eIF－4B、eIF－4A 和 eIF－3 等因子复合物的组装。这类复合物是因子识别 mRNA 的活性形式，有益于加快翻译的起始速率。与 eIF－4E 的磷酸化作用的重要相比，γ－亚单位的磷酸化意义仍不太清楚。

2．其他因子磷酸化对翻译的激活作用

起始因子 eIF－4B 是与 mRNA 相结合，对其他与 mRNA 作用的起始因子 eIF－4A 和 eIF－4B 的活性起协调或激活作用。在促有丝分裂剂作用下，eIF－4B 随 eIF－4F 和核糖体蛋白 S6 一起，在细胞内被磷酸化。eIF－4B 的磷酸化作用是随机的，他有 8 个或更多的丝氨酸位点可供磷酸化反应。eIF－4B 可被 PK－C 以及哺乳细胞或蟾蜍的 S6 激酶磷酸化。目前不未发现专一对 eIF－4B 磷酸的激酶。

3．eIF－2 的磷酸化对翻译的抑制作用

eIF－2 是由分子量分别为 36 000、38 000 和 55 000 Da 的 3 个亚基组成。eIF－4F 能特异地与 Met-tRNAi 结合，并且这种结合必须有 GTP 参加。eIF－2 的磷酸化会引起翻译起始作用受阻，从而抑制蛋白质的生物合成。

二、酵母 GCN4 mRNA 翻译的调控

酿酒酵母（*S. cerevisiae*）GCN4 是一个转录因子，它与启动元件 TGACTC 相结合，参与了包括编码氨基酸生物合成酶系各基因在内的表达调节作用。大多数酵母基因的任一种氨基酸饥饿，均会增强 GCN4 因子的合成。GCN4 的增多完全是由于编码 GCN4 的 mRNA翻译作用的激活，该激活是由于酵母细胞翻译起始因子 eIF－2 的磷酸作用。前文明确提出了 eIF－2 的磷酸化是翻译起始作用的负调控因子，而对酵母 GCN4 mRNA 则有增加蛋白质翻译量的功能。其原因是：在氨基酸饥饿时，空载的 tRNA 被堆积，这些tRNA 可以作用于 GCN2 蛋白激酶肽链上的类似 His-tRNA 合成酶样结构，活化的激酶使 eIF－2因子 α 亚单位的第 51 位上的丝氨酸磷酸化，磷酸化的 eIF－2 抑制由 eIF－2B 参与的 GTP 及 eIF－2 的再循环，而启动了 GCN4 mRNA 的翻译作用。

图 8－13 示意的 GCN4 mRNA 及其 5′先导序列中，GCN4 的开放可读框（GCN4-ORF）起始于第五个 AUG 密码，在此之前的各个 AUG 形成了 4 个短的上游可读框（upstream open reading frame，uORF），从 5′-帽子到第一个 uORF 的 AUG 之间的序列可以使核糖体在第一个 AUG 起始的 uORF1 启动翻译。在正常非饥饿情况下，活化型的 eIF－2 的水平是高的，在 uORF1 翻译之后，释放下来的核糖体可以重新往下游滑动，并在其他任一个 uORF 再启动翻译，而 GCN4 mRNA 的 ORF 的翻译被排除了；然而饥饿时，磷酸化的 eIF－2α 增加，减少了活化形式的 eIF－2 再循环，使在 uORF1 翻译之后，释放下来的核糖体滑动时，穿过其他 uORF，而结合于 GCN4 mRNA 起始 AUG 位置，重新启动 GCN4 ORF 的翻译。

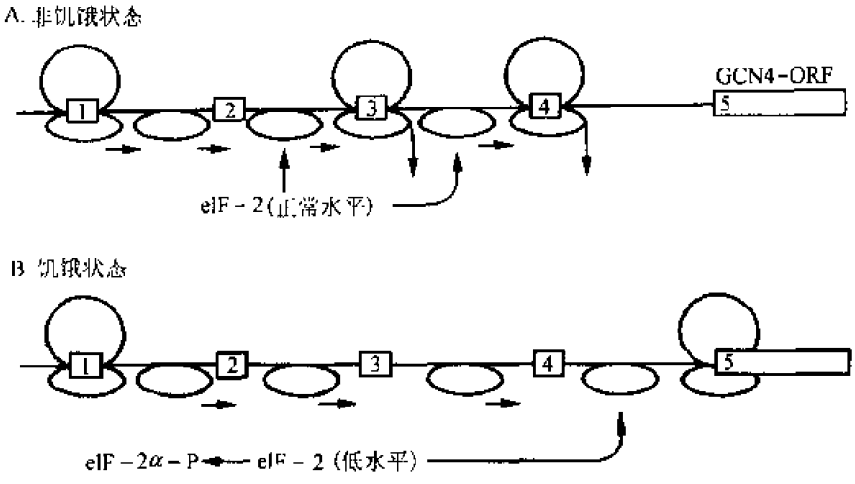


图 8-13 酵母 GCN4 mRNA 的翻译模式

1~4 表示短的 uORF;5 表示起始子 AUG 的 GCN4 mRNA 可读框。

三、转铁蛋白 mRNA 翻译的调控

mRNA 分子结构的本身与其翻译调控有着密切关系,而其调控作用往往通过其结构元件与相应蛋白质因子的相互作用而实现。在研究翻译调控中,这种 RNA 和蛋白质可逆性相互反应的典型模式是转铁蛋白的翻译控制。转铁蛋白(ferritin)和红细胞的 5-氨基乙酰丙酸合成酶(5-aminolevulinate synthase, ALAS)两种 mRNA 的翻译调控与铁离子密切有关。其 mRNA 靠近 5′帽子部位的一个茎环式结构,被称为铁离子应答元件(iron-responsive element,IRE)。胞浆内一种分子量为 98 kDa 的蛋白质能特异地结合在 mRNA 的 IRE 上,这个蛋白质称为铁离子调节因子(iron regulatory factor, IRF),也曾有人称它为铁离子应答元件结合蛋白(iron-responsive element binding protein, IRE-BP)、FRP 或 P90。当 IRF 结合到 mRNA 的 IRE 元件上后,就封闭了 mRNA 的翻译作用。当细胞铁离子的水平低时,IRF 对 IRE 有高的亲和力,从而导致转铁蛋白和 eALAS 的合成抑制。增加铁离子的获取,可减少 IRF 对 IRE 的亲和力,可去除对翻译的抑制作用。因此,这种调节作用的基础是依赖于铁离子多少控制阻抑蛋白对 mRNA 结合位点的控制。

(欧阳红生　李慎涛)

参 考 文 献

[1] 沈琪琲等. 北京:高等教育出版社,1998.14
[2] 汪玉松等.北京:中国农业科技出版社,1999.350
[3] Lewin B. New York:Oxford University Press and Cell Press,1997. 335
[4] Yamamoto K R et al. Ann Rev Genet,1985. 19:209
[5] Gehring W J et al. Cell,1994. 78:221
[6] Boyes J et al. Cell,1991. 64:1123
[7] Chailet J R et al. Cell,1991. 66:77

第九章　遗传重组

第一节　概　　述[1-14]

基因组在时间上和空间上都不是固定的，它们结合了为遗传所要求的稳定性和变异所要求的可变性，基因组的可变性和稳定性之间必须维持一个恰到好处的平衡，这样才能使生物体得以生存并能世代相传，繁衍不息。

行有性繁殖的双倍体机体在减数分裂时，同源染色体彼此独立分配，在配子中存在着染色体的不同组合。当雄性和雌性配子融合时，便产生一个基因组不同于该物种其他成员的个体。这种情况的发生是通过彼此不同的染色体再分配形成的，是一个简单但却十分重要的重组机制，但在该过程中没有 DNA 间遗传物质的物理交换。在不同染色体上的基因重新分配，不需要有物理交换，但它需要染色体的遗传差异。如果一个个体的染色体与同种的其他成员是一样的，独立分配和有性繁殖则只产生遗传上一样的个体。染色体的遗传差异主要由两种机制产生，一种是突变，一种是遗传重组。所谓遗传重组指的是遗传物质的重排，其共有的特征是 DNA 双螺旋之间的遗传物质发生交换，因发生的机制不同分为几种类型。遗传重组系统的存在，确保了物种中代与代之间基因组的重排，从而形成一个物种个体之间的遗传差异。在减数分裂时染色体的独立分配、基因突变、遗传重组共同作用降低了一个物种个体间的遗传相似性。

遗传重组不仅仅发生于代与代之间，一个个体的基因组也可以发生重排。这能产生基因表达的改变，在一个个体的细胞之间产生遗传基因的多样性。这些过程不能通过染色体重新分配产生，但能通过重组机制产生。

重组不只是在减数分裂和体细胞核基因中发生，也在线粒体基因间和叶绿体基因间发生。重组也见于噬菌体整合过程中和转座子转座等过程中。地球上所有的机体可能都有一些重组机制，这表明重组对物种的生存是十分重要的。生物要生存便得适应，要适应便得有变异，而变异的来源便是突变和重组。重组可使有利和不利的基因分离，并作为一个单元在新的组合中被检验。它提供了一种方式，使有利的等位基因得以保存，使不利的基因被清除。通过重组获得基因的新的重组体允许该物种更快的适应环境并加速进化的进程，同时重组对 DNA 损伤的修复也起重要作用。

突变和重组提供进化起始物质的两个过程，突变引起的遗传改变是引起蛋白质中氨基酸序列的变化，该变化引起表型的改变，通过自然选择发生作用。相类似地，重组提供一个基因组结构的变化，引起表型的变化，也通过自然选择发挥作用。重组反应在体细胞的分化和损伤 DNA 的修复中也起重要作用。重组在生物学的中心位置反映在遗传学中它的中心作用。突变和重组是遗传学家的重要工具。

根据不同的机制可将重组分成 4 类，同源性重组（homologous recombination）、位点特异性重组（site-specific recombination）、转座（ transposition）和异常重组（illegitimate recom-

bination)，有人曾把后三种称为非同源性重组，或只把第四种称为非同源性重组，但现在认为，即使异常重组，有时也需要微小的同源性存在，所以现在把重组分成上述4种基本类型(图9-1)。

(1)同源性重组　同源性重组是指发生在同源DNA序列之间的重组。同源性重组主要是利用DNA序列的同源性识别重组对象，蛋白质(例如，大肠杆菌RecA蛋白)可以促进识别，但是提供特异性识别的是碱基序列。

(2)位点特异性重组　位点特异性重组(site-specific recombination)发生在特殊的序列对之间，当一种蛋白与一个识别位点相结合并催化同另一个位点进行重组，该位点也被同一蛋白质或同一重组系统的另一种成分识别。在重组对之间可以有同源性，因为它们两个部位结合同样的蛋白质或该反应一个中间步骤要求同源性。但蛋白质和DNA，蛋白质和蛋白质之间的作用是关键。

(3)转座重组　转座指一段DNA序列插入到另一个DNA序列中而不依赖序列的同源性，转座可使一定的DNA片段从一个染色体的一个部位转移到另一部位。它也是涉及到蛋白质与DNA和蛋白质与蛋白质的相互作用，但在这里，只是重组对的一个被催化转座的蛋白质识别。该蛋白质正确地识别转座子末端，然后同一个相对无限的靶部位反应。

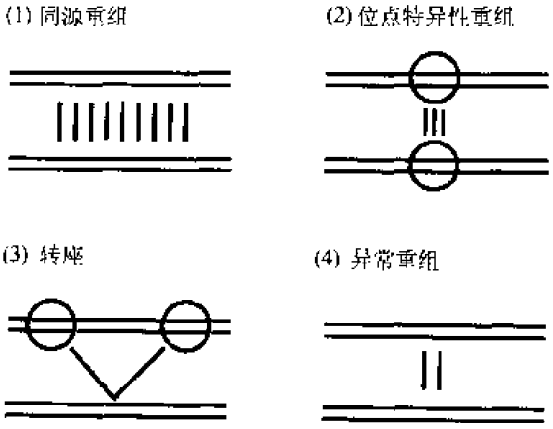


图9-1　遗传重组的4种类型
(1)同源性重组涉及到重组对之间DNA序列同源性的识别。同源性重组要求同源性的最低长度因不同的机体而异，一般在25～300 bp的长度。(2)位点特异性重组通过结合在特定靶序列的蛋白质所介导并催化的在这些部位的重组。重组位点的识别涉及到蛋白质与DNA和蛋白质与蛋白质的相互作用。(3)转座发生在一个限定的DNA序列(称为转座元件)，一种蛋白质能识别转座元件的末端和靶序列，该蛋白质由转座子编码。(4)异常重组发生在短的DNA序列，低于同源性重组所要求的长度。这些序列可以有一些同源性但也可没有同源性。垂直和斜线代表重组对的识别，圆环代表蛋白质识别特异DNA序列。

(4)异常重组　异常重组分为两类，末端连接和链滑动。在这两种反应中的特征是重组对中很少或没有序列同源性，因此，异常重组有时被称为非同源性重组。但因在异常重组中微小同源性常有一种作用，所以“非同源性重组”一词易引起误导。

重组反应分类为同源性重组、位点特异性重组、转座和异常重组是在该领域里产生一种逻辑分类的标准方式(也便于更好地理解在重组过程中的不同机制)。但是应当考虑到许多重组反应利用不止一种机制进行，例如：Tn*10*的转座涉及到转座和同源性重组；免疫球蛋白VDJ连接涉及到位点特异性识别和一种类似异常重组的末端连接过程，酵母接合型的转换既具有同源性重组和位点特异性重组的特征，也具有复制转座过程的特征等等。另外在有关的重组机制没有十分清楚之前，分类也是很困难的。本书将较详细介绍同源重组、转座(包括反转座)、位点特异性重组、异常重组和免疫球蛋白多样性的产生以及由重组等原因引起的DNA序列的重排和扩增。

重组是遗传学的基石，能够定位基因的彼此位置，然后再用这些定位信息对实验机体的染色体进行操作和理解，是现代生物学的心脏。不仅仅同源性重组被广泛地应用于该目的，而且其他的重组机制也是如此。转座元件可以将基因引入到染色体，或灭活基因。

位点特异性重组系统引入特异重组反应是理想的 DNA 序列。异常重组能被用于整合任何 DNA 序列到一个基因组中。通过同源性重组将 DNA 序列引入真核基因组已经为研究高等生物开辟了一条新途径。一个含有靶基因的细胞产生一个完整的转基因机体的能力,探究在整个机体中一个特殊基因的作用是可能的。

通过遗传图已经可能研究遗传性疾病,疾病基因座的鉴定已经能够诊断许多疾病。最后,重组技术对基因治疗新技术的发展正在做出贡献,基因治疗能针对遗传病本身引起的原因进行治疗。野生型基因引入到受突变基因影响的组织中以补偿失去的或不正确的信息。这些技术和方法在几年后可望得到迅速发展。

第二节　同源性重组[1~28]

一、同源性重组是重要而复杂的过程

在细胞里,两个巨大的 DNA 分子连接起来并交换序列是个十分复杂的过程。同源性重组常常只发生在两个 DNA 分子的同源区。这样这些区域必须排列起来,以便它们能被割裂并再连接。在重组点的两侧必须改变它们彼此的构象。这一复杂的过程涉及到多种功能。

同源性重组是两个 DNA 双螺旋之间的反应。它的最主要特征是相关的酶可以利用任何一对同源序列为底物(虽然一些序列要优于另一些)。整个基因组的各个部分重组频率并不保持恒定,它要受全局及局部因素的影响。总体的频率在卵细胞和精细胞中可能是不同的,在人类,女性重组的频率是男性的两倍。在基因组中其频率也取决于染色体的结构。例如,在异染色质附近遗传物质的交换要受到抑制。

同源性重组发生在 DNA 的同源序列之间,真核生物减数分裂时的染色单体之间的交换,某些低等真核生物及细菌的转化、转导、接合,噬菌体的重组等都属于同源性重组这一类型。

二、进行同源性重组的基本条件

(1)*在交换区具有相同或相似的序列*　同源性重组的显著特征是在发生交换的 DNA 的两个区域的核苷酸序列必须是相同或很相似的。同一物种的不同个体中,DNA 分子核苷酸序列常常是几乎一样的。两个同源染色体的 DNA 分子同样区域中序列一般是相同的。同源性重组常常只发生在两个 DNA 分子中的相同部位。两个 DNA 分子中不同部位间的重组有时也会发生,因为在 DNA 分子中,同样或类似的序列有时可见于多处。这种类型的重组称为异位重组,它可引起 DNA 序列的缺失、复制、倒位和其他 DNA 的重排等。细胞已经产生一种特殊的机制以防止异位重组的发生。

(2)*双链 DNA 分子之间互补碱基进行配对*　我们知道只有一种机制可以使核酸以序列为基础而相互识别:这就是单链之间的互补。两个 DNA 分子的链之间互补的碱基配对确保重组只发生在同样的基因座之间,也就是 DNA 分子的相同的部位。在该部位两个双链 DNA 分子通过链之间互补的碱基配对被维系在一起称为联会,所有的同源

性重组必须涉及到联会的形成。

(3)重组酶　使两条 DNA 双螺旋分子间断裂、修复、连接和重组体的释放是重组的中心环节。在重组中两个 DNA 分子之间的序列发生交换。因此,每个分子的链必须首先断裂,然后进行修复和连接。重组完成后,重组分子的解离和释放等过程均是在酶催化下进行的,参与重组反应的酶保证重组的顺利进行。在原核生物中这些酶研究得较为清楚。

(4)异源双链区的形成　同源性重组时,在两个 DNA 分子之间互补碱基配对的区域称为异源双链区。一个亲代 DNA 双螺旋分子与另一个亲代 DNA 双螺旋分子通过一段异源 DNA 双螺旋共价相互连接。因为这些区域的链来自不同的 DNA 分子。也就是由两个亲代 DNA 分子的链形成异源双链 DNA。

三、同源性重组的分子模型

1. Holliday 双链侵入模型

第一个被广泛接受的重组模型是 1964 年由 Robin Holliday 提出的 Holliday 模型(图 9-2)。根据这个模型,在将发生重组的两个 DNA 分子中,通过在两个 DNA 分子的同一部位两个单链的断裂而引发重组。然后两个断裂的单链的游离末端彼此交换,每一条链同另一分子的互补序列配对形成两个异源双链。然后末端彼此连接产生十字样的结构称为 Holliday 连接体(Holliday junction),它由两个双链分子通过它们交换的链连在一起。Holliday连接体是所有重组模型的核心(图 9-3)。

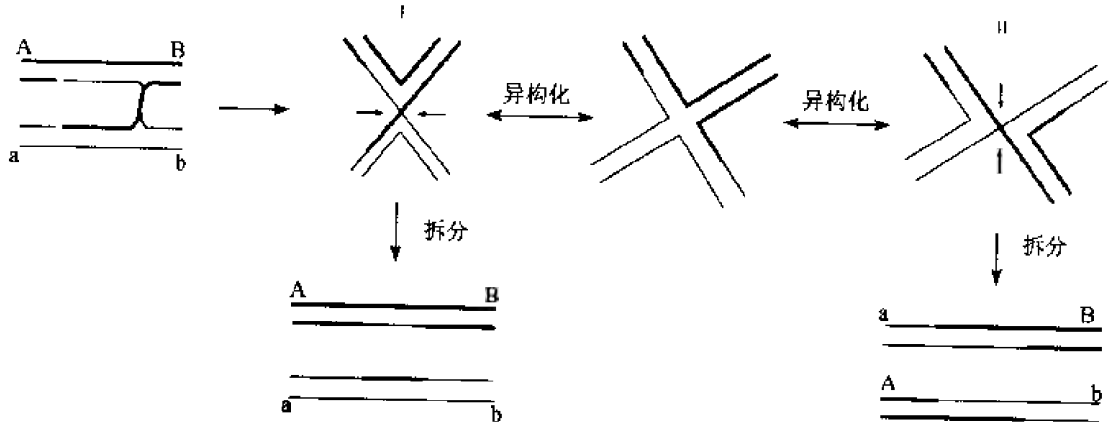
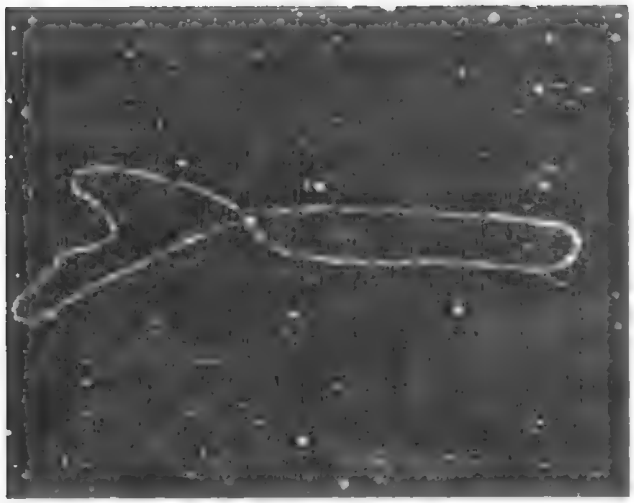


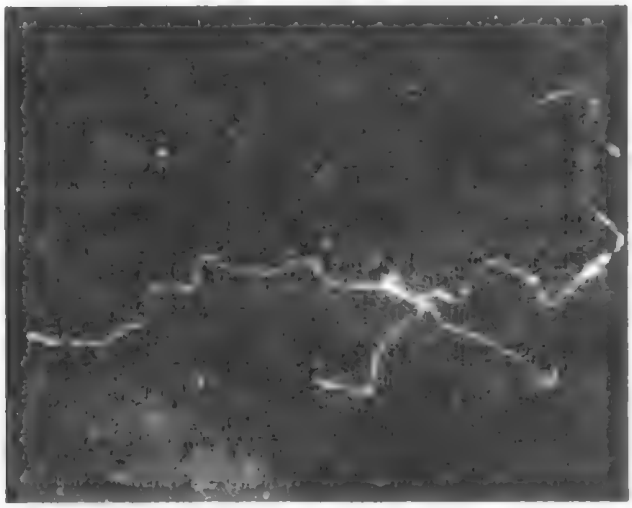
图 9-2　Holliday 遗传重组模型

每个 DNA 分子的一条链在同样位置被切开,然后同另一个分子的互补链配对形成异源双链。然后链被连接形成 Holliday 连接体。这种 DNA 结构在 I 型和 II 型之间可以异构化。拆分后使 Holliday 连接体分离,子代 DNA 取决于 Holliday 连接体的构象,两侧的遗传标记 A、B、a 和 b 将被重组或仍然处于原来的构象,子代 DNA 分子将含有异源双链区部分。

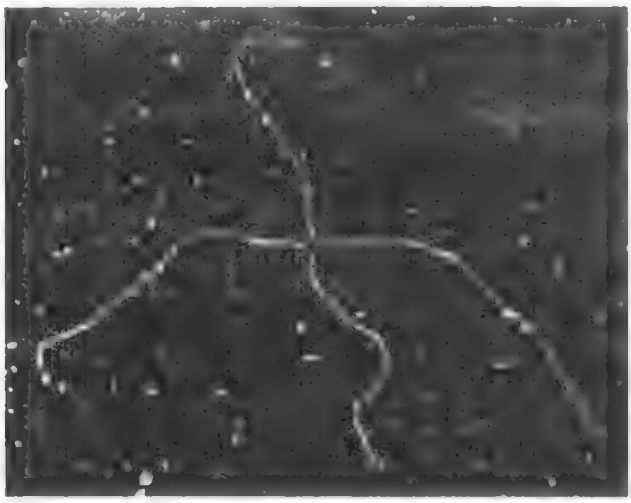
一旦 Holliday 连接体形成后,它能进行重排从而改变链的彼此关系。这种重排称为异构化,因为在此过程中没有键的割裂。如图 9-2 所示,第一种异构化引起交换链处在没有交换的构象 I 的位置。第二种异构化产生构象 II,在构象 II 中,两个双链 DNA 分子



(1)



(2)



(3)

图 9-3　细菌质粒存在 Holliday 连接体的证据(电镜照片)

(1)由大肠杆菌提取质粒,一小部分呈现 8 字形分子结构,这是由两个环形质粒分子以 Holliday 连接体的形式融合在一起形成。(2)应用在质粒上存在一个切点的限制酶切割质粒,8 字形中间体经酶切后产生一个十字形结构。(3)DNA 经部分变性后展开,在 Holliday 连接体处,可见 4 条链(H. Potter and D. Dressler)。

的末端处在彼此重组的构象中。Holliday 连接体能由一种类型转变为另一种类型而在碱基之间没有任何氢键的断裂。由一种构象转变为另一种构象不需要能量，所以能很快发生，每种构象存在的概率各占 50%。

一旦形成 Holliday 连接体后，就能被拆分。是否发生重组依赖于拆分时 Holliday 连接体的构象。如果连接体是在构象 I 拆分，两侧的 DNA 序列将不发生重组。但每一个双螺旋皆含一段由于交换事件而留下的遗传信息发生变化的 DNA 片段。然而如果 Holliday 连接体是以构象 II 形式存在，当它拆分时，DNA 序列将被交换，象在两侧遗传标记所指出的那样(图9－2)。

Holliday 连接体也能通过碱基之间氢键的断裂和再连接而发生左右移动。这个过程称为支链迁移(branch migration)见图 9－4。当交互连接移动时，同样数目的氢键断裂和再形成，不需要能量。但是在没有能量的实验中，氢键断裂的不够迅速，所以不能有效地进行支链迁移。特异的 ATP 水解蛋白似乎是在支链迁移中所需要的。如图 9－4 所示，在 Holliday 连接体中支链的迁移能增加异源区长度。如果一个异源双链扩展到含有不同序列的区域，将发生错配，这可能导致基因转换。

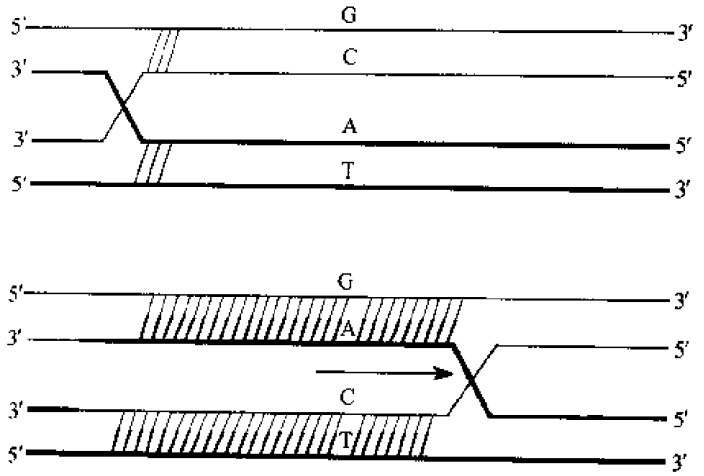


图 9－4 Holliday 连接体的迁移

两条 DNA 分子之间形成的交互点可以沿 DNA 向任意方向移动，这一过程叫链迁移。链迁移改变了交叉点的位置和异源双链的长度。但 Holliday 连接体的基本结构不变。通过破坏链前边的将 DNA 维系在一起的氢键，然后再形成氢键。两个 DNA 分子间连接部位的迁移，扩展了配对区(也就是异源双链区)。异源双链区用斜线表示，因为一个 DNA 分子在该区有一个突变，在异源双链区形成两个错配部位 GA 和 CT。

Holliday 模型被称为双链侵入模型，因为由每一个 DNA 分子的一条链侵入到另一个 DNA 分子，它解释了在重组时，两个 DNA 分子的异源双链是如何形成的。然而，这种模型存在的一个问题是两个 DNA 分子必须同时几乎在同一部位被切断而引发重组。但是当碱基被藏在双链 DNA 螺旋的内部，而不能随意地与另一个 DNA 分子配对时，两个相似的 DNA 分子在它们被切断之前是如何为配对而排列在一起的呢？如果两个 DNA 分子不被排列在一起，它们如何能在确切的同样部位被切断呢？为了回答这个问题，Holliday 认为在 DNA 分子上存在某些位点，这些位点能被特殊的引发重组的酶切断。然而，还没有足够的证据证明这些位点的存在，而重组似乎多少有些随机地发生在整个 DNA 分子的任何部位。

尽管有这些问题，Holliday 双链侵入模型已被作为一个标准模型。几乎所有的重组模型都涉及到 Holliday 连接体和支链迁移。它们之间的差别主要存在早期阶段，在 Holliday 连接体形成之前。

2. 单链侵入模型

有一种方式可以克服 Holliday 模型中存在的问题，就是单链侵入模型，是对 Holliday

模型的修改(图 9-5)。在该模型中,两个 DNA 分子中的一个单链在随机部位被切断,然后暴露的末端侵入到另一个双链 DNA 分子,直到发现它的互补序列。如果它发现这样一个序列,它将替代一条与它相同的链。然后 DNA 聚合酶用留下的那条链做为模板,将充填上侵入链所留下的缺口。被替代的链被降解,它的残留末端与另一分子中新合成的 DNA 链相连接形成 Holliday 连接体。一旦 Holliday 连接体形成,像 Holliday 模型一样,产生异构化然后拆分。

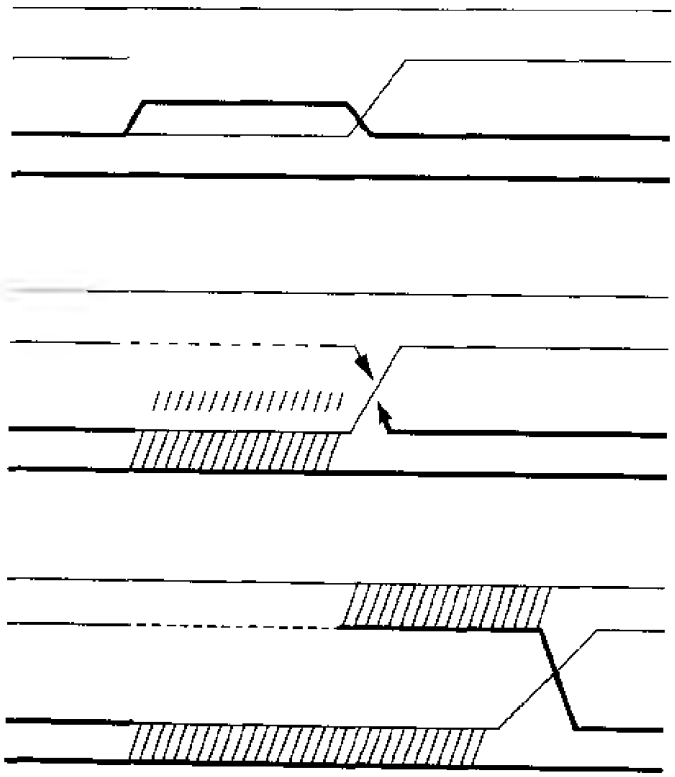


图 9-5 单链侵入模型

单链侵入模型。在两个 DNA 分子中的一个单链发生断裂,游离出一个单链末端侵入到另一个 DNA 分子中。被切割的 DNA 留下的缺口由 DNA 聚合酶进行修复(虚线)。在另一个 DNA 分子中被替代的链降解,而两个末端被连接(见箭头)。开始,一个异源双链(用斜线代表)将只在两个 DNA 分子中的一个形成,支链迁移将在另一个 DNA 分子上产生另一个异源双链。像在 Holliday 模型一样,异构化使两侧的 DNA 分子重组。

通过这个模型,异源双链首先只在两个 DNA 分子中的一个形成。然而一旦 Holliday 连接体形成,支链迁移能在另一个 DNA 分子上产生异源双链。这就解释了两个 DNA 分子中异源双链是如何形成的。

3. 双链断裂修复模型

在 DNA 分子中,通过双链断裂引发重组看起来是不可能的。如果在 DNA 分子两条链中一条链发生断裂,另一条链仍然将分子维系到一起,然而如果双链发生断裂,DNA 分子的两部分可能被分开,这可能是一个致死的过程。这样双链断裂模型首先被排除。然而,现在似乎是很明确了,两个 DNA 分子之中的一个双链发生断裂可引发重组,至少在某些情况下是这样。

当前的重组模型认为，遗传信息的交换是由双链的断裂而引发的。DNA 双链断裂引发重组第一个证据来自酵母的遗传实验。后来证明，通过双链断裂引发重组是个常见的机制。如细菌、噬菌体和低等真核生物的归巢 DNA 内切酶（homing DNA endonucleases）可因产生双链断裂而引发重组。

双链断裂重组模型认为，参与重组的一对 DNA 分子之一的两条链被核酸内切酶切断，然后在核酸外切酶作用下扩展为一个缺口，并在（一种或几种）核酸外切酶的作用下产生 3′单链黏性末端，此两个 3′游离末端之一侵入到另一个双螺旋的同源区，置换“供体”双螺旋的一个单链而形成一段异源双链 DNA，并同时产生一个 D 环（D-loop）。此 D 环由于利用 3′游离末端为引物，在 DNA 聚合酶作用下的修补合成而扩展。最终 D 环的长度变得与“受体”染色体的缺口长度相当。当突出的单链到达缺口的另一端时，互补的两条单链退火。此时在缺口的两侧各有一段异源双链 DNA，并且此缺口被 D 环单链 DNA 所占据。

缺口处双链的完整性可以被以缺口 3′端为起始的修补合成来恢复。总的来说，缺口是被两次单链 DNA 的合成而修复的（图 9－6）。

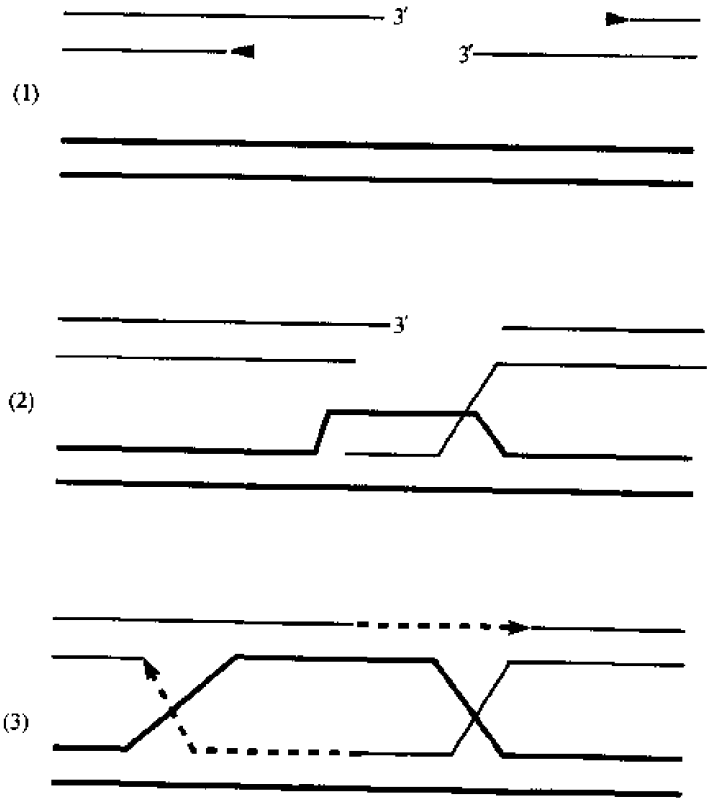


图 9－6　双链断裂修复模型

（1）两个 DNA 分子中一个 DNA 分子双链断裂，箭头指出 5′端发生降解。（2）3′端侵入到另一 DNA 分子中并置换出同源链。（3）DNA 聚合酶以 3′端为引物进行修复直到和 5′末端连接（黑色箭头）。两个 Holliday 连接体形成。

支链迁移使这一结构转换成一个具有两个重组连接体的分子。也就是有两种 Holliday 连接体形成。是否发生重组依赖于在拆分时两个 Holliday 连接体是处于哪种构象。如果两者处于相同的构象 I 或 II（图 9－2），当它们被拆分时，则不发生交换，这样就不发

生重组。且每一个双螺旋皆含一段异源双链。然而如果两个 Holliday 连接体是处在不同的构象时，拆分后将发生重组。

双螺旋断裂后，参与交换的区间的两侧皆有杂合 DNA 的形成。在此两个杂合双螺旋之间为断裂的缺口，并且此缺口为"供体"DNA 序列所填充。因此异源双螺旋序列的排布是非对称的，一个双螺旋分子的部分序列被另一双螺旋分子的相应序列所替换（这就是引发重组的染色单体被称为"受体"的原因）。

双链断裂模式并没有降低形成异源双链 DNA 的重要性，并且仍为两个双螺旋分子间相互作用唯一似乎合理的方式。

在前两个模型中，在重组过程的任意阶段都没有遗传信息的丢失。但在双链断裂模型中，起始断裂之后，紧接着就是遗传信息的丢失。在恢复这些信息的过程中，任何错误都可能是致命的。但另一方面，正是通过另一双螺旋而重新合成丢失信息的这一能力提供了细胞的主要安全屏障。

四、联会复合体

真核细胞染色体的联会是在分子水平上最难解释的一个阶段。染色体复制后进入减数分裂过程，同源染色体配对形成联会复合体。许多年来人们一直认为它有可能代表重组过程中 DNA 交换的一个预备阶段。最近的观点认为联会复合体是重组的结果而不是其发生的原因。但是不论哪种情况，我们皆不清楚联会复合体的结构与 DNA 分子相互接触间的确切关系。

在这一阶段，每个染色体皆呈现为以两条侧成分作为边界的染色质的集合体，而且两个侧成分被一个细微但密集的中央成分所分离。此三股平行而密集的线状结构位于同一平面并围绕其轴线而扭曲。在此结构中同源染色体间的距离大于 200 nm（DNA 分子直径为 2.0 nm）。因此理解此复合体作用的一个主要难点在于，虽然它使同源染色体排列到一起，但远没有做到使它们之间相互接触。

联会复合体两侧之间唯一可见的连接物为见于真菌和昆虫的球状或圆柱状结构，它们横跨此复合体并被称为小结（nodes）或重组小结（recombination nodules），它们产生的频率和分布的情况与染色体交叉（chiasmata）相同。它们的名称表明人们希望证明它们即是重组的发生位点，但它们是否是重组发生的位点，有待证明。

近期的实验表明在果蝇（*Drosophila*）或酵母中，所有阻止染色体配对的突变同时也阻止重组的发生。在少数几个系统中，重组事件可以在分子和细胞水平上相比较，酿酒酵母（*S. cerevisiae*）减数分裂过程的研究工作已经取得了进展，图 9-7 为重组在分子和细胞水平上的对应的时间表。

已有很好的证据表明，在酵母中双链断裂引发同源性重组和位点特异性重组。在一个序列为另一个序列所取代的接合型的变换中，首先涉及到双链断裂。双链断裂也发生在减数分裂早期的特殊位点上，这些位点被称为重组的热点。重组的频率在热点的一侧或两侧以梯度形式衰减。热点即为重组起始点，梯度反映出重组过程可能是由此而扩展开的。

我们现在可以在分子水平上解释这一观点。如图 9-6 所示，双链断裂产生两个平头

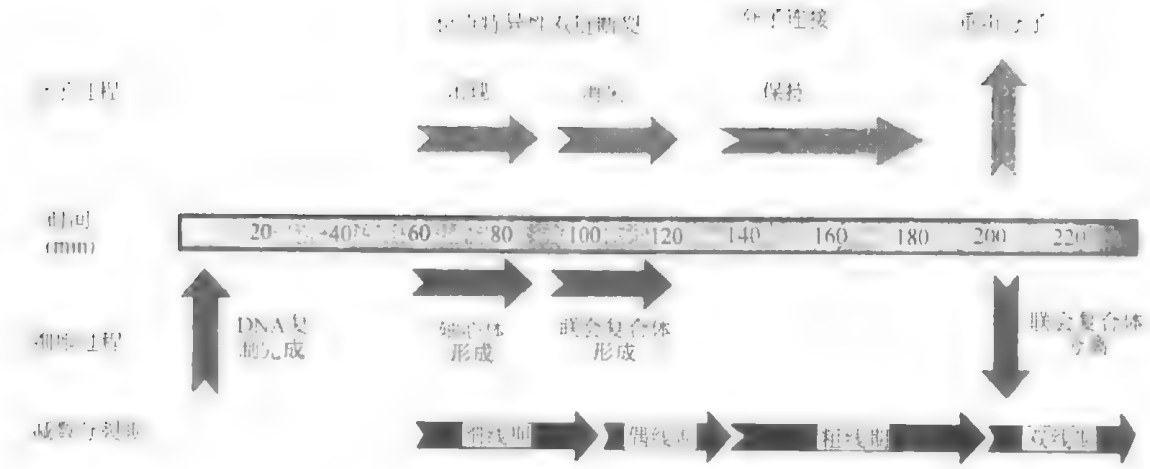


图 9-7　重组在分子水平和细胞水平上对应的时间表

的 DNA 末端，两个平头末端很快转换为 3′单链黏性末端。一个不能使平头末端转换为单链黏性末端的酵母突变体（*rad 50*）不能进行重组。这表明双链断裂是重组所必须的。梯度的产生则是由于产生单链区的可能性随其与断裂点之间的距离增加而降低。

双链断裂的出现和随后消失大约经过 60 分钟。双链断裂消失后不久即产生第一个交联分子，并且，一般认为，此交联分子是重组的中间体。这些过程的发生顺序表明双链断裂、配对反应和重组子结构的形成是在染色体的同一位点连续发生的。

双链断裂发生在轴心体（axial element）形成期间，并且随着配对的染色体转换成联会复合体的过程而消失。这一相对的时间顺序表明联会复合体的形成是由双链断裂引发重组和随后转换成重组中间体的结果。实验观测到的 *rad 50* 突变种不能使轴心体转换成联会复合体的事实支持这种观点。这就说明传统的，认为在减数分裂中，联会复合体的存在表明染色体只有在配对后才能发生重组的观点是不正确的。

确定重组是否发生在联会阶段是很困难的，因为重组是通过对减数分裂后重组子的产生而评估的。但是，在酵母，重组子的产生可直接根据含有诊断性限制性酶切位点（diagnostic restriction sites）的 DNA 分子的产生进行评估。实验表明，在粗线期的末期即有重组子产生，这清楚地表明，重组过程起始于联会复合体形成之前，重组过程的完成在联会复合体形成之后。

所以，联会复合体的形成是在引发重组的双链断裂之后，并一直保持到重组分子的形成。这表明在染色体配对后联会复合体的形成是染色体重组的一个结果，并且是随后的减数分裂阶段所必需的。一个联会复合体形成的酵母突变体（*zip1*）能去除交换干扰（一个重组过程能抑制附近另外一个重组发生的能力）。这表明联会复合体的形成可能在重组位点被引发，而且此复合体沿染色体的扩展本身抑制进一步重组事件的发生。在这种情况下，重组事件可能并不完全是被同时引发的，并且联会复合体有可能在重组引发时开始形成并在重组尚未结束时完成。这一关于联会复合体抑制重组的看法和早期的关于它对重组起主要作用的假说形成了鲜明的对照。

从分子水平来看，我们可以推测当同源染色体相互识别之后（依某种未知方式），双链的断裂产生单链末端，并且此末端开始了搜寻互补序列的过程。搜寻的结果可能是两个

双螺旋的同源区发生连接并使它们相互间足够接近并引发联会复合体的形成。我们仍不清楚当重组体被观察到之前，在粗线期发生了什么。也许重组的一些随后步骤如链交换的扩展，DNA 的合成和重组体的释放在这一期间发生。

在减数分裂的下一阶段（双线期），联会复合体消失，染色体间的交叉变得可见。这些连接点已被假定为遗传信息发生了交换的表征，但是在分子水平上的特性并不清楚。它们有可能代表完全交换的残留物或代表遗传信息发生交换后尚未解离的染色体同源部分的连接。在减数分裂的随后阶段，这些连接点迁移到染色体的末端。这种可变性表明这些交叉点代表重组的残留物而不是重组的中间体。

重组发生在减数分裂的染色体的分立的位点上，但是我们还不能把这些位点与已观察到的分立的结构单位（重组小结和交叉点）相关联。但是在酵母中，对一些存在于这些分立位点上的蛋白质的鉴定有助于在分子水平上提供了解这些分立结构形成的一些线索。这些蛋白质包括 MSH4（与一种参与细菌不配对修复的蛋白质有关）、DmC1 和 Rad51C（大肠杆菌 RecA 蛋白的类似物）。这些蛋白质在重组中的确切作用尚不清楚。

重组是一个受控制的过程，只有一小部分经相互作用而最终完成交换。通常情况下，每对同源染色体只获得 1～2 个交换，同源染色体不发生交换的可能性特别低（<0.1%）。这种过程可能是单一交换控制的结果，因为在某些突变体中，交换的非随机性通常被破坏。并且，重组过程是完成减数分裂所必须的。由于检验系统的存在，如果没有发生重组，减数分裂会受到抑制。当重组完成后，此种抑制则消除。这一安全系统确保在重组未发生之前，细胞不会发生分裂。

五、基因转换

同源性重组时有异源双链 DNA 形成，处于异源双链区两侧的基因在 Holliday 连接体拆分时，可以发生交互重组，也可不发生交互重组，对于位于异源双链区内的基因或遗传标记，不论拆分以何种方式进行都可能在这些部位发生遗传学变化。异源双链 DNA 双螺旋的参与解释了等位基因间重组的特性。等位基因间的重组对异源双链模型的发展起到了推动作用。同源性重组时，在重组位点染色体发生断裂和重新结合，并由此产生一个互补对。在单一基因附近，异源双链 DNA 的形成本身和重组有关。可以通过子囊菌（*Ascomycetes*）来研究独立的重组事件，这是由于在子囊真菌中每一次减数分裂的产物被包裹于一个大的细胞中（被称为子囊），更为有利的是由减数分裂而产生的 4 个单倍体核排成一列。实际上，这 4 个核在产生之后又发生了一次有丝分裂，因此共有 8 个单倍体核排成一列。图 9-8 所示，这 8 个单倍体核的每一个皆有效地代表了由减数分裂所产生的 4 个染色体的 8 条 DNA 单链的遗传特征。

在大多数孢子中，一个杂合子经减数分裂，每一个等位基因应产生 4 个拷贝。但在一些孢子中这一比例则显示异常。这一现象可被解释为在等位基因不同的区域内由于形成异源双链 DNA 并在等位基因有差别的区域进行校正所造成的。此图所示为由于重组的发生使 4 条经减数分裂的染色体的两个含一段杂合 DNA，这是由双链断裂而引发重组的结果。

假设两个等位基因只因单一位点的突变而存在差异，当经过链的交换而产生杂合

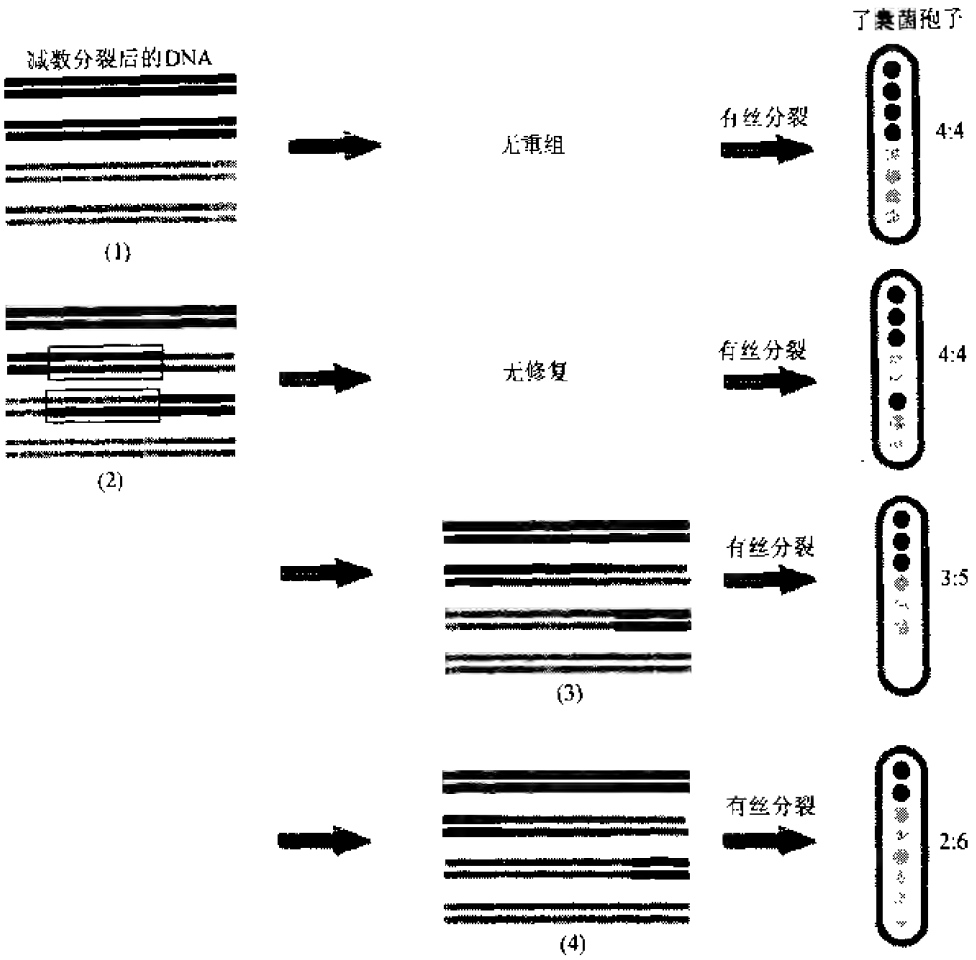


图 9-8　子囊菌孢子形成过程中的基因转换

(1)无重组;(2)有重组但无修复,有丝分裂后分离;(3)一条异源双链被修复出现 3:5 的分离比例;(4)两条异源双链按同一方向进行矫正,出现 2:6 的分离比例。

DNA 后,这一杂合 DNA 双链则会在突变位点处不匹配,每一个 DNA 链携带不同的遗传信息。如果此序列不发生变化,双链分离并随后复制,则每一个链产生一个 DNA 双螺旋,并且自身的遗传信息得以保留,这一过程被称做减数分裂后分离(postmeiotic segregation),因为它反映了在减数分裂之后 DNA 双链的分离。其重要性在于它直接证明了在重组的 DNA 等位基因中存在异源双链 DNA。

在考查等位基因的重组过程中所发现的另一现象为:等位基因的比例与初始的 4:4 有所偏离。这一现象被称做基因转换(gene conversion)。它所描述的是一条染色单体与另一条染色单体发生了非交互重组。

基因转换是 DNA 分子间链的交换所造成的,在分子水平上,这种序列的变化有可能由如下二种原因引起的:

如图 9-6 双链断裂模型所示,一个 DNA 双螺旋可以做为遗传信息的供体,通过链的交换和缺口的填充而直接取代受体中相应的序列;作为交换过程的一部分,当一个双螺旋的一条 DNA 链与另一个双螺旋的互补区发生配对时,则产生异源双链 DNA。当修复系统识别不匹配的碱基后,有可能对一条链的不配对碱基进行切除并代之以配对的碱基,而

使互补性得以恢复。这种通过对异源双链区内不配对碱基的修复而进行的基因校正过程称为基因转换。这一事件使代表一个等位基因的DNA链转换为另一等位基因序列。

基因转换并不依赖于交互重组的发生，但却与其相关。很大一部分异常的子囊显示在等位基因内发生转换的位点两侧的标记间发生了重组。如果此种异常的比例是重组引起的结果（如图9-2或图9-6所示的那样以产生和不产生重组大致相同的概率拆分重组体），则这一结果正与期望的相符合，这表明真菌染色体引发交换的频率大致是远距离基因重组频率所得数值的两倍。

当在分子水平上考查重组时，可以发现种种倾向性。基因能以相同的概率向任一方向转换，或由于等位基因的特殊作用而产生对某一方向的倾向性。重组的梯度也许会偏离热点。我们现在知道热点为双链断裂的引发点，并且重组频率的梯度与断裂点的间隙的扩展并转化成长的单链末端有关。

基因簇各成员的序列可以提供关于基因转换程度的一些信息。通常，一个重组的产物会相互分离，因此不能进行DNA的序列分析。但是，当一个染色体含两个相互关联的非等位的基因时，它们可以通过一个"不等量交换"过程而相互重组。两个非等位基因间可以形成一个异源双链。基因转换过程可有效地使一个非等位基因转化为另一个基因序列。

在一个染色体上存在一个基因的多个拷贝的现象提供了追踪这些过程的线索。例如，如果一个基因的一部分形成杂合双螺旋并发生了基因转换，则这一部分的序列与另一个基因序列可能会完全相同或十分接近，而另一部分的序列则会显示较多的差异。已知的序列分析表明，基因转换过程可能会延伸相当长的距离，甚至达到几千个碱基之多。

第三节　大肠杆菌重组的分子基础[29~56]

和真核生物相比，大肠杆菌重组的分子基础已经了解的比较清楚，和重组有关的酶已得到鉴定，它们在重组中的作用已基本阐明。因此，为了研究DNA分子序列间相互交换过程的特性，我们必须转向细菌系统。在这里识别反应是重组机制的不可分割的重要部分，并且只涉及到DNA分子的特定区域而不是完整的染色体。但在分子水平上重组事件发生的先后顺序是相似的：从一个已断裂的分子中产生的单链与其对应的双螺旋发生作用，配对区间扩展，重组中间体形成到核酸内切酶解离此两个双螺旋分子。现已知每一阶段所需的酶，虽然它们大概只代表重组所需的某些组成部分。由于含 *rec*⁻ 基因突变种的出现，细菌中参与重组的酶已得到鉴别（表9-1）。*rec*⁻ 突变种的表型为这些突变种缺乏进行一般重组的能力。

表9-1　大肠杆菌遗传重组所需要的蛋白质

蛋白质	活　　性
RecA	DNA链交换，DNA复性，赖DNA的ATP酶，赖DNA和ATP的蛋白酶
RecBCD（外切核酸酶V）	解旋酶，赖ATP的dsDNA和ssDNA外切酶，ATP促进的ssDNA内切酶，识别 *Chi* 序列
RecBC	DNA解旋酶

续表

蛋白质	活　　性
RecE[外切核酸酶Ⅷ(*sbc*A*)]	5′→3′方向的 dsDNA 外切酶
RecF	结合 ssDNA、dsDNA、ATP
RecG	Holliday 连接体的分支移动，解旋酶
RecJ	5′→3′方向的 ssDNA 外切酶
RecN	有 ATP 结合位点的共有序列，确切功能不明
RecO	与 RecR，可能还有 RecF 相互作用
RecQ	DNA 解旋酶
RecR	与 RecO，可能还有 RecF 相互作用
RecT	DNA 复性
RuvA	与 Holliday 十字形四臂连接体结合，与 RuvB 相互作用
RuvB	Holliday 连接体的分支移动，DNA 解旋酶，与 RuvA 互作
RuvC	Holliday 连接体切割，结合四臂连接体
SbcB(外切核酸酶 I)	3′→5′方向外切 ssDNA，脱氧核糖磷酸二酯酶
SbcCD	赖 ATP 的 dsDNA 外切
SSB	结合 ssDNA
DNA 拓扑异构酶 I	ω 蛋白、I 型拓扑异构酶
DNA 促旋酶	DNA 促旋酶，II 型拓扑异构酶
DNA 连接酶	DNA 连接酶
DNA 聚合酶 I	DNA 聚合酶，5′→3′外切，3′→5′外切
解旋酶 II	DNA 解旋酶
解旋酶Ⅳ	DNA 解旋酶

sbcA 突变是调节突变，它有激活 *recE* 基因的功能。

(引自 S.C. Kowalczykowski et al.，Microbiol. Rev.，1994. 58：401～465)

一、*Chi* 位点和 RecBCD 核酸酶

在细菌中通常没有大量的双螺旋 DNA 的交换，并且在原核生物中可能有许多引发重组的方式。在一些情况下，DNA 可以游离单链 3′末端的形式存在，即 DNA 可以单链的形式被提供(例如在细菌接合过程中)，单链也可产生于辐射损伤，单链末端也可由噬菌体基因组经滚环式复制而产生。但是，在涉及到两个双螺旋分子的情况下，按照真核细胞在减数分裂中所遇到的重组规则，则必须产生单链区和 3′末端。

由于激发重组的某些热点的存在，现已发现了一种产生适宜末端的机制。研究发现在 λ 噬菌体的一些突变种中，单一碱基对的改变即可产生激发重组的一些称为 *Chi* 的位点。这些位点引导我们发现了一些与重组相关联的蛋白质的作用。

这些位点皆含有一个恒定的非对称的 8 bp 序列：

5′　GCTGGTGG　3′

3′　CGACCACC　5′

这一 *Chi* 序列天然存在于大肠杆菌的 DNA 中，并且在大约每 5～10 kb 长的序列中即出现一次。在野生型 λ 噬菌体 DNA 和一些其他的遗传单位中，此序列并不存在，这说

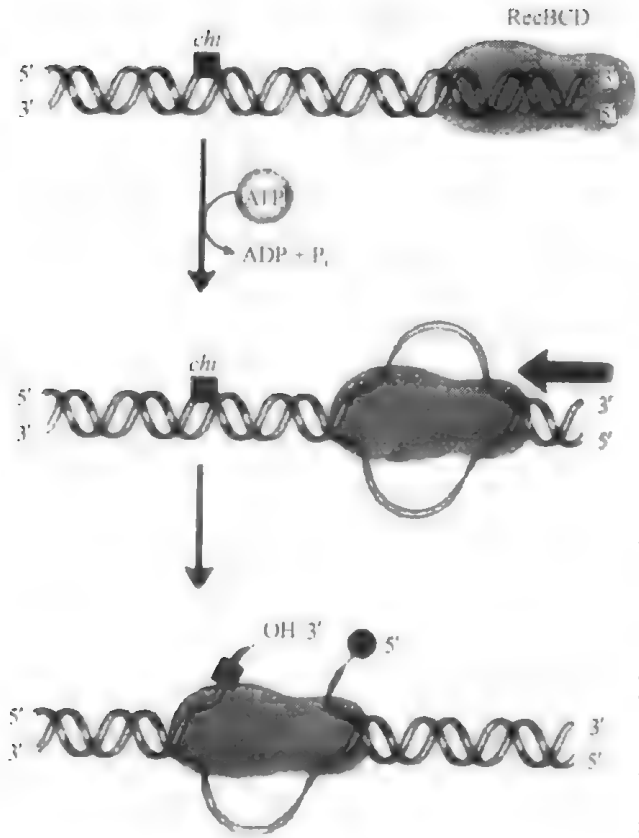


图 9-9 Rec BCD 蛋白的作用模型

Rec BCD 蛋白作用于双链 DNA 的末端，通过该蛋白质的解旋酶和单链内切核酸酶活性使 DNA 解旋，并在单链内切核酸酶活性的作用下在 *Chi* 位点的右侧 4～6 个核苷酸处切断单链 DNA 链。继而由 RecA 蛋白促进同源重组。

明它并不是重组所必须的。

Chi 序列在其附近(大约 10 kb 序列长度内)激发重组。一个 *Chi* 位点可以被在其单一侧(上述序列的右侧)相距几个 kb 的双链断裂所激活，这一取向依赖性表明重组装置(recombination apparatus)与 DNA 的断端相连，并只能沿双螺旋的一个方向移动。

Chi 位点是一个由基因 *rec BCD* 编码的 Rec BCD 酶作用的靶部位，Rec BCD 酶是一种多功能酶，具有几种不同的活性。它是一个强有力的降解 DNA 的核酸酶，早期被检定为活性核酸外切酶 V。它具有解旋酶的活性，在 SSB 的存在下，它能够解旋双链 DNA。并且它具有腺苷三磷酸酶(ATPase)的活性和序列特异性单链内切酶的活性，它在重组中的作用大概是提供一个含游离 3′末端的单链区。

图 9 9 显示这些反应是如何在一个含 *Chi* 序列的 DNA 底物上协调进行的。当 Rec BCD 在 *Chi* 序列右侧与 DNA 相结合后，它沿 DNA 移动并使 DNA 解旋。当它到达 *Chi* 位点时，暂停移动，并在 *Chi* 序列右侧 4 到 6 个碱基处切断一条单链(上面的链)。由于对 *Chi* 序列的识别使 Rec D 亚单位脱落或失活，结果使整个酶失去核酸酶的活性。但它仍做为一个解旋酶而起作用。

Rec BCD 所介导的解旋和切割可以被用来产生末端并引发异源双链的形成。RecA 能够利用由 Rec BCD 在 *Chi* 位点附近切割所释放的单链的 3′末端，与同源的双螺旋序列起反应，并产生一个交联分子。

二、RecA 蛋白和 Holliday 连接体的形成

RecA 具有两种相当不同的活性：它可在 SOS 反应中激活蛋白酶；也可以促进单链 DNA 与其在双链 DNA 分子中互补的 DNA 链的碱基相配对。

RecA 需要单链 DNA 和 ATP 才能激活蛋白酶，在它操作 DNA 分子时也许要此两种底物。现在仍不十分清楚在体内 RecA 的此两种酶活性与重组的关系，但其相关联的几个反应提供了关于重组机制的一些有价值的启示。

RecA 的 DNA 操作酶活性可以使一个单链 DNA 与双螺旋中的同源序列发生置换，

这一反应被称为单链摄取(single-strand uptake)或单链同化(single-strand assimilation)。这一反应可以发生在几种构象不同的DNA分子类型之间,并且包含以下三种限制条件:参与反应的一个DNA分子必须含一段单链区;参与反应的一个DNA分子必须含一个游离的3′末端;单链区和3′末端必须位于两个DNA分子互补区之内。

当一条线型单链DNA侵入双螺旋后,此双螺旋中的与单链同源的序列则被置换。当"供体"或"受体"DNA任意之一为环型时,这一反应最易被跟踪。反应沿着将被置换的DNA链的5′→3′方向进行,这就表明,这一反应涉及到的参与交换的DNA链至少含一个游离的3′末端。

单链的同化与重组的引发具有潜在的相关联性。所有的重组模型皆须一个双螺旋的一条或二条单链与另一双螺旋DNA分子相互作用的中间体的存在。RecA可以催化这一阶段的反应。

RecA能与单链或双螺旋DNA相结合而形成长的丝状结构的能力形成了RecA激发支链迁移的一种可能机制。这种丝状物呈螺旋形结构,每一螺旋含6个RecA分子单元,并且其深沟中含DNA分子。定量的关系是每个RecA分子与三个核苷酸(或碱基对)相结合。此型DNA为相应的B型DNA双螺旋长度的1.5倍,每一螺旋含18.6个碱基。当DNA双螺旋与RecA相结合时,它与RecA在小沟相结合,大沟则留出来用于与另一个DNA分子反应。

两个DNA分子间的反应发生在这些丝状结构之内。当一个单链被双螺旋所吸收前,第一步是单链与RecA相结合而形成丝状结构,其后与双螺旋相结合,大概形成一种三链结构(图9-10)。

链的交换需要游离的3′末端。反应发生在丝状结构之内,并且RecA保持与最初的单链相结合,这样在反应完成时RecA与杂合DNA相结合。在反应过程中大量的ATP被水解,ATP有可能通过使RecA发生变构而起作用。当与ATP结合后,RecA的DNA结合位点对DNA具有高亲和力。这是与DNA相结合和配对反应所需要的。ATP的水解使结合位点变成低亲和性,这是释放杂合DNA所需要的。

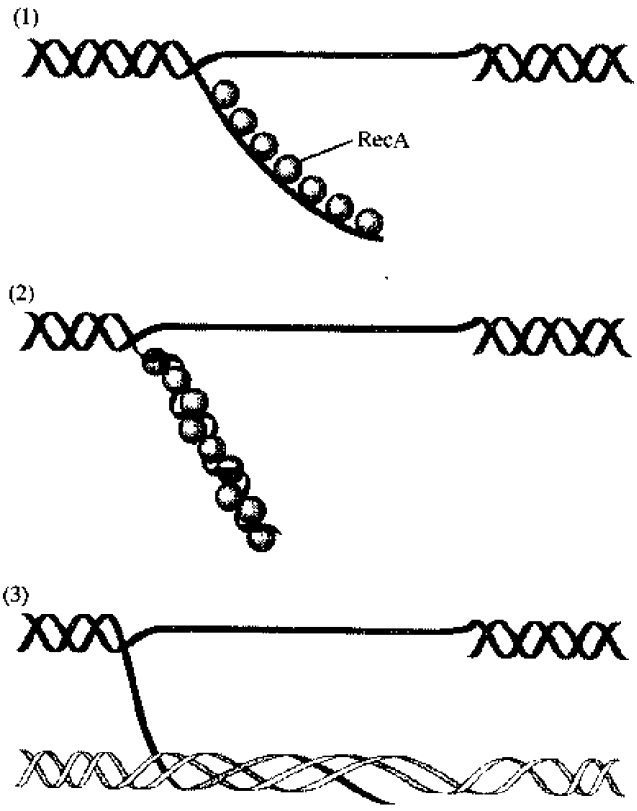


图9-10 RecA蛋白的作用方式
两个同源的DNA分子之间通过RecA蛋白的作用发生联会。在(1)、(2)和(3)中,RecA蛋白和单链DNA相结合并迫使单链进入双螺旋结构形成三链结构。促进同源性重组。

我们可以将RecA所催化的单链与双螺旋间的反应分成二个阶段:

RecA 分子在单链 DNA 上相聚合的缓慢的前联会期；单链 DNA 与其在双螺旋中相互补的序列的快速的配对，产生一个杂合双螺旋连接点；一个缓慢的对双螺旋的一条链的置换反应，产生一个长的杂合 DNA 双螺旋。

SSB 的存在可激发此反应，因为它可以确保底物缺乏二级结构。现在并不清楚 SSB 和RecA是如何对同一段 DNA 起作用的。和 RecA 一样，SSB 所需要的量也是有一定比例的，这表明它在链的吸收中所起的作用大概是协同式的，与 DNA 分子相结合而形成与丝状结构相关联的结构。

当一个单链 DNA 分子与双螺旋反应时，在重组区此双螺旋分子发生解旋。最初所形成的异源双链 DNA 甚至有可能不以通常的双螺旋形式存在，有可能这两个单链只以并列的方式相连接。一个这种类型的区域被称为平行汇接(paranemic joint)(可与经典的在双螺旋中链的相互缠绕的关系(plectonemic)相比较)。平行汇接并不稳定，反应的继续进行要求其转化成双螺旋形式，这一反应相当于解除负的超螺旋，并且有可能需要酶来产生短暂的链断裂而使其可以相互旋绕(rotate)，从而解决加旋/解旋的难题。

我们所讨论的这些反应只代表重组过程的一部分：单链对双链的侵入。RecA 也可使两个双螺旋分子发生相互作用，条件是其中之一含有一个至少 50 个碱基长的单链区。这一单链区可以一个线型分子的末端存在，或者一个环型分子的一个裂口形式存在。

一个部分双螺旋分子与一个完全双螺旋分子间的反应可以导致单链的交换。吸收在线型分子的一端开始，侵入的单链以通常的方式置换其同源链，当反应进行到两个分子的双螺旋区时，侵入的单链与其初始配对的链相分离，并且新产生的单链与被置换的单链相配对。

在这一阶段，分子的结构与重组连接是不可区分的。在体外，通过 RecA 的作用可产生 Holliday 连接体，这表明这种酶可以介导链的相互交换。我们对与 RecA 相结合的四链中间体的几何结构所知不多，但假定两个双螺旋分子以交换反应所要求的相适合的方式并排而列。

三、Ruv 和 RecG 蛋白和 Holliday 连接体的拆分

在大肠杆菌中，有一组由三个基因编码与随后的重组相关连的蛋白。*ruv A* 和 *ruv B* 基因的产物促进异源双链的形成。Ruv A 蛋白可以识别 Holliday 连接体的结构，Ruv B 是一个腺苷三磷酸酶并可能提供支链迁移的动力。这一复合物的结构如图 9－11 所示，Ruv A 在交叉点处与 DNA 所有的 4 条链相结合，在交叉点的上游，两个 Ruv B 六聚体环状结构分别与每个双螺旋相结合。Ruv AB 蛋白复合体可以使支链以 10～20 bp/秒的速度迁移。RecG 解旋酶也具有相似的活性。在 Ruv AB 解旋酶作用过程中，要置换与 DNA 相结合的 RecG。Ruv AB 和 RecG 皆具有拆分 Holliday 连接体的作用，并使重组得以完成的活性，如果二者皆为突变体，则大肠杆菌完全失去重组能力。

第三个基因 *ruv C*，编码一种能特异性地识别 Holliday 连接体的核酸内切酶，此酶可以在体外切断此种交叉而拆分重组中间体。一个含 4 个碱基的序列提供了 *Ruv C* 拆分 Holliday 连接体的热点。这一序列为非对称结构，因此可能决定在拆分过程中哪一对 DNA 链被切断，从而决定发生的是补钉型重组(patch recombinantion)(没有整体的重

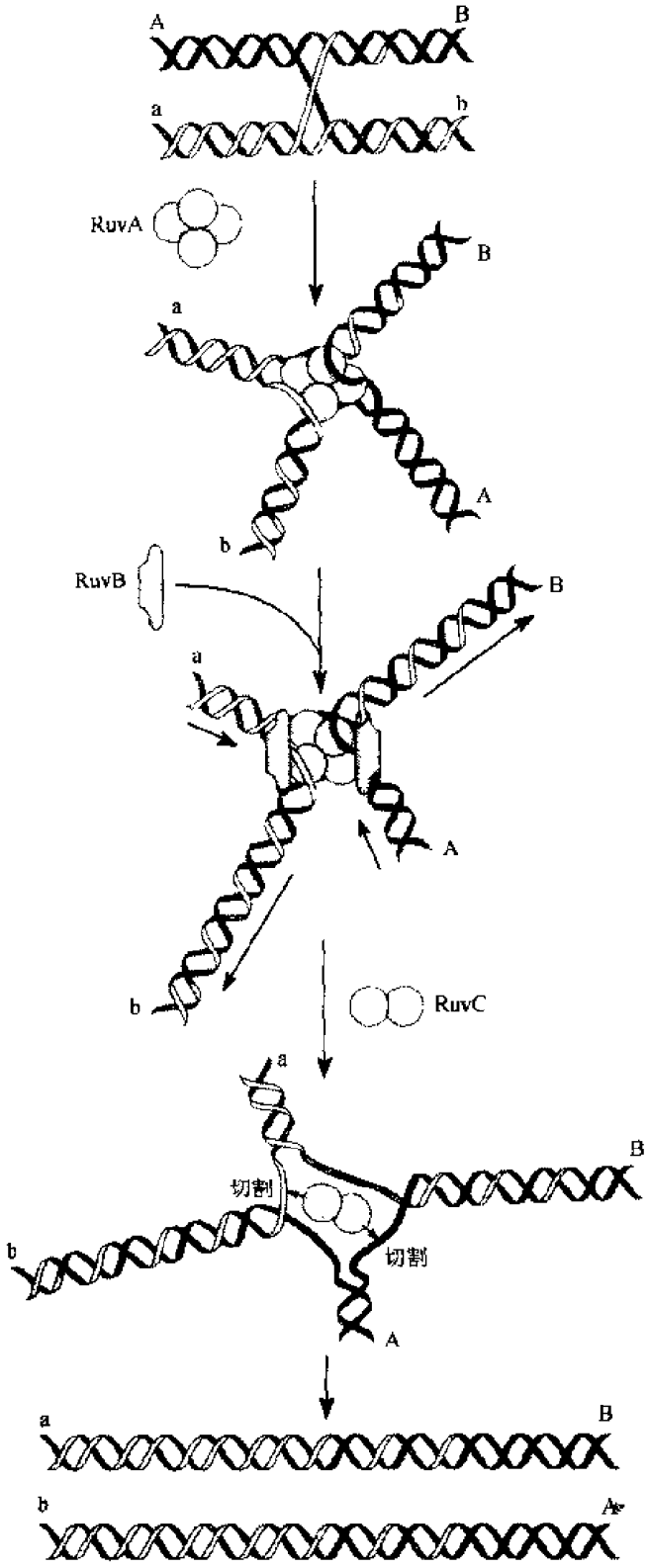


图 9-11　Ruv 蛋白作用机制的模型

RuvA 蛋白与 Holliday 连接体结合，然后 RuvB 与 RuvA 蛋白结合，RuvC 切割 Holliday 中间体的两条链，拆分后形成两条 DNA 分子。

组)，还是剪接型重组(splice recombination)(两侧的重组)。

我们现在可以从每个蛋白质所起的作用角度来解释重组的每个过程。利用重组可从一个双螺旋获取物质来修复另一双螺旋的缺口。利用这些结论来解释真核生物重组的主要不足在于在细菌中这些过程主要涉及到一个DNA片段与整个染色体的作用，并且是由DNA损伤引发的修补反应，其与发生在减数分裂过程中基因组的重组并不完全等同。然而，涉及操作DNA的分子水平上的活动是相同的。

RecA的同系物普遍存在于原核生物中，并且相关的蛋白在真核生物中也有发现。在酿酒酵母中，存在两个基因，*DMC1* 和 *rads1*，它们所编码的蛋白与RecA相关。这些基因的突变体产生相似的表现型，双链断裂的累积和不能形成正常的联会复合体。这进一步支持以下观点：DNA双螺旋间链的交换与联会复合体的形成有关，并且表明染色体联会与细菌链的同化反应可能是相关的。但RecA在真核生物中的同系物并不形成丝状结构，因此在真核生物中反应的机制很可能是不同的。

第四节　DNA的拓扑学[57~61]

对DNA的拓扑操作是其所有的功能(重组、复制，也许包括转录和组成高级结构)的一个中心环节。所有与双链DNA有关的合成要求两条链相互分离。但是，这两条链并不是简单的并排而列，它们是相互缠绕的。其相互分离要求二者在三维空间相互旋转。

我们可以设想由于DNA游离末端的存在，两条单链似乎可以绕着双螺旋的轴旋转而解旋。但是当考虑到双螺旋的长度时，这种解旋要求DNA链绕轴进行大量的摆动，由于细胞体积的限制，这种解旋似乎不大可能。

如果在游离末端放置一个能控制旋转的装置则可以获得相似的结果。但是，这种效果必须被传送过相当的距离，这同样会涉及到过长的DNA链的旋转。DNA实际上表现为没有游离末端的封闭结构，这就在原则上排除了以上这些模型，并由此而产生了严重的扑拓上的困难。考虑在一个分子中末端不能自由旋转的两个链分开时所产生的效果：当两个相互缠绕的链在一端被分离时，这两条链在远端的相互缠绕则会相应增加。因此复制交叉点的移动则会在其前端产生更多的正超螺旋，并对其继续向前移动产生难以克服的阻力。

这个困难可以用在一个链上引入瞬间切口的方式来解决。一个内部的游离末端可使被切断的链绕另一完整的链旋转，切点可以随后被连接。切断和连接反应的每一次重复则会释放一个超螺旋。

我们知到封闭的DNA分子可以用连接数 L(在三维空间一条链绕另一条的缠绕次数)来描述。具有相同序列的封闭的DNA分子可能具有不同的连接数，这表明分子具有不同程度的超螺旋。除了连接数不同而其余皆相同的DNA分子被称为拓扑异构体。

连接数由缠绕数 W 与扭转数 T 之和组成，因此连接数的变化被表示为 $\triangle L = \triangle W + \triangle T$。一种比较松散的表述方式为：连接数的变化为双螺旋在空间绕轴缠绕数的变化($\triangle W$，相当于超螺旋)与双螺旋自身螺旋数的变化($\triangle T$)之和。一个自由的DNA分子其 W 和 T 值可以自由调整，并且连接数的变化很可能被表述为超螺旋(W)的变化。

任何连接数的变化皆需要至少一条链的断裂。利用此游离末端，一条链可以绕另一

条旋转，之后断裂点被重新接合。此种反应使一种拓扑异构体转化为另一种。DNA 拓扑异构酶催化此种类型的转化。一些拓扑异构酶可以从 DNA 分子中去除负的超螺旋；另一些可以除去正和负的超螺旋；一些可引入负的超螺旋。

根据不同的工作原理，拓扑异构酶可以分为两类：第一类酶会使单链 DNA 产生短暂断裂，第二类酶则会引入双链 DNA 的短暂断裂。拓扑异构酶可以是序列选择性的，也可是非选择性的。被充分鉴定的第一类拓扑异构酶是大肠杆菌 *topA* 基因的产物，其功能为使含很高负的超螺旋的 DNA 松弛。此酶对正的超螺旋 DNA 不起作用。如此酶发生突变则会使超螺旋水平增加（并可能影响转录）。

除了使含负超螺旋的 DNA 松弛外，此酶还与单链 DNA 相互作用。这大概与负的超螺旋有类似性，因为负的超螺旋倾向于稳定单链区，而这种单链区可能是酶结合的底物。当大肠杆菌拓扑异构酶 I 与 DNA 相结合时，与 DNA 形成一个稳定的复合物，DNA 的一条链被割裂并且其 5′的磷酸末端与酶的酪氨酸残基共价结合。这显示了酶的一种可能作用机制：它使 DNA 的一个磷酸二酯键转移到酶上，并随后使断裂的 DNA 链重新接合。

化学键从核酸到蛋白的转移解释了为什么此酶可以在无须提供能量的情况下起作用的原因。这是因为在反应中没有不可逆键的水解。在这些转移反应中能量被保存了下来。

拓扑异构酶 I 的一个作用是，酶首先使双螺旋 DNA 变成两条相互分离的 DNA 链并在此部位与 DNA 结合，然后使一条链断裂，另一条链则通过断裂间隙被拉到另一侧，最后断裂点被重新接合。真核生物的第一类拓扑异构酶的性质相似，但它们能使正的和负的超螺旋松弛。

反应只用一步来改变连接数，完整的单链每一次通过断裂点时，$\triangle L$ 的值就增加 1。对一个自由的超螺旋分子，由于 W 和 T 的相互可换性，连接数的变化可以通过 $\triangle W = +1$ 来表示，也就是，减少（$\triangle L = +1$）次负的超螺旋。

旋转过程是等价的，所加的限制是酶催化的每一次反应只有一次单链通过缺口区。（相反的，在一个超螺旋的分子中引入一个断裂点会使单链自由旋转并直到所有张力皆被释放为止）。

第一类拓扑异构酶也可以使单链 DNA 的一段穿过另一段。这种单链通过反应可在 DNA 分子中引入结，并可以使两个环型分子像链子一样连接起来。我们并不了解这些反应在体内的作用（如果有的话）。

第二类拓扑异构酶通常可以使正和负的超螺旋松弛。反应需要 ATP，大概每一个催化反应需要水解一个 ATP 分子。反应使一个 DNA 双螺旋分子双链断裂，并使另一双螺旋由此通过。这种双链转移的一个正式效果是使连接数以 2 的倍数而变化。拓扑异构酶 II 也可以被用来引入或解离相连接的双螺旋环和有节的分子。

反应大概表明酶对双螺旋 DNA 分子的非特异性识别，酶与任意两个双螺旋片段相结合并使其相互交叉。ATP 的水解大概提供酶的构型变化以动力，并通过此变化使一个 DNA 双螺旋通过另一个 DNA 双螺旋的断裂点。由于 DNA 超螺旋分子的拓扑特征，这种过程会使正或负的超螺旋被去除。

促旋酶引入 DNA 分子负的超螺旋。细菌 DNA 促旋酶属于第二类拓扑异构酶，它能够引入负的超螺旋到松弛并封闭的环型分子。DNA 促旋酶和环型 DNA 双螺旋结合后，

以催化的方式逐步引入超螺旋;它可以连续地将超螺旋引入同一个分子。一个 DNA 促旋酶分子每分钟引入大约 100 个超螺旋。

超螺旋的 DNA 分子比松弛状态含有较高的自由能,完成转化所需的能量由 ATP 的水解所提供。在没有 ATP 存在的情况下,促旋酶可以使负的(不是正的)超螺旋松弛,但是反应速率要比引入超螺旋慢十多倍。

大肠杆菌 DNA 促旋酶是一个含两种亚单位的四聚体,每一种亚单位皆是抗生素的作用目标(常用的有作用于 GyrA 的为萘啶酮酸,作用于 GyrB 的为新生霉素)。这些药物抑制复制,表明 DNA 促旋酶是进行 DNA 合成所必需的。对这些药品产生抗药性的突变体的研究鉴定了编码这些亚单位的基因座。

促旋酶在四聚体的外周与底物 DNA 相结合。促旋酶可保护大约 140 bp 长度的 DNA 不被微球菌核酸酶(micrococcal nuclease)所降解。酶与呈交叉状构型的 DNA 分子(相当于一个正的超螺旋)相结合,这会在 DNA 的非结合区引入一个补偿性的负的超螺旋,随后酶在正的超螺旋的交叉点上切断一个 DNA 双螺旋,让另一个双螺旋经过此断点,并接合断点。

反应直接改变了超螺旋从 +1 到 −1 的符号。因此连接数的改变为 $\triangle L = -2$,进一步证实了所有涉及双链通过的过程必会使连接数改变为 2 的倍数的要求。

促旋酶随后释放(新变负的)超螺旋片段。这使负螺旋在 DNA 上可重新分布(以 T 或/和 W 的变化方式),随后循环重新开始。同样的拓扑操作可产生结和环型连结。

当已经变换的超螺旋被释放后,促旋酶的构型发生了变化。为了使酶可重新催化另一个超螺旋循环,它的初始构型必须重新恢复。这一过程被称为酶的转换(enzyme turnover),并被假设为由 ATP 的水解所驱动。这是因为当用不可水解的 ATP 相似物替代 ATP 时,促旋酶只能对每一个底物引入一次转化(−2 超螺旋)。

因此它的超螺旋反应不需要 ATP,但需 ATP 进行第二次循环。新生霉素阻止 ATP 与亚单位 B 的结合,因此干扰促旋酶的依靠 ATP 的反应。不依靠 ATP 的松弛反应可被萘啶酮酸所抑制。这表明 A 亚单位参与了断裂和重新结合反应。用萘啶酮酸作用酶,可允许 DNA 以横跨 DNA 的交错剪切所产生的片段的形式被回收。所有的片段末端皆含一个游离的 3′-OH 基,和一个 5′的由 4 个碱基所组成的与 A 亚单位共价相连的延伸,共价的相连使磷酸键中的能量得以保存。它可被用来驱动接合反应,并解释了促旋酶不依靠 ATP 催化松弛反应的原因。切点的选择具有相当的特异性,大约为每 100 bp 发生一次反应。

第五节　位点特异性重组[62~77]

位点特异性重组发生在特殊序列对之间,这一重组型式最早是在 λ 噬菌体的遗传学研究中发现的,当噬菌体侵入大肠杆菌细胞后,λDNA 有两种存在型式:裂解状态和溶源状态。在裂解状态下,λDNA 在被感染的细菌中以独立的环形分子结构存在。在溶源状态下,噬菌体 DNA 则是细菌染色体的一部分(称为原噬菌体,prophage)。两种类型间的转换是通过位点特异性重组实现的。

一、λ噬菌体 DNA 的整合与切除

为了进入溶源状态，游离的 λDNA 必须整合(integrate)到细菌 DNA 中去；而为了从溶源状态向裂解状态转化，原噬菌体 DNA 则必须从细菌染色体 DNA 上切除(excise)。整合和切除均通过细菌 DNA 和 λDNA 上特定位点之间的重组而实现。这些特定位点叫做附着点(attachment site，att)。在细菌遗传学上，细菌染色体上的附着点，被称为 att^{λ}。这一位点是从能阻止λDNA整合的突变体上鉴定的。在溶源性菌株上，这一位点被原噬菌体 λDNA 所占据。当从大肠杆菌的染色体上去除 att^{λ} 位点后，感染性 λ 噬菌体可以整合其 DNA 在染色体的其他部位而建立溶源状态。但其反应效率仅相当于存在 att^{λ} 时频率的 0.1% 以下。这一低效的整合发生在第二附着点(secondary attachment site)上。此点与 *att* 序列存在相似性。

为了描述整合切除反应，细菌的附着点(att^{λ})称为 *attB*，含有 *B*、*O* 和 *B′* 三个序列(*BOB′*)。λ 噬菌体的附着位点称 *attP*，由 *P*、*O* 和 *P′* 三个序列组成。其中 *O* 序列是 *attB* 和 *attP* 所共有的，序列完全一致，称核心序列(core sequence)。是位点特异性重组发生的地方。两侧的 *B*、*B′* 和 *P*、*P′* 称为臂，其序列各不相同。由于线状的 λDNA 在侵入细胞后不久就首尾连接成环状，所以在 *att* 处的相互重组导致了整个 λDNA 整合进细菌染色体 DNA 中。在整合状态下，λ 噬菌体呈线状序列。原噬菌体位于两个新的 *att* 位点之间，这两个位点是重组的产物，不同于原来的 *attB* 和 *attP*。原噬菌体左边的位点是 *attL*，由序列 *BOP′* 组成，右边的是 *attR*，由序列 *POB′* 组成。上述位点的差异说明，整合和切除反应所需要的序列对不相同。整合要求对 *attP* 和 *attB* 之间进行识别，但切除要求对 *attL* 和 *attR* 进行识别。因此位点特异性重组的方向性特征由重组位点的特征所控制。

虽然重组过程是可逆的，但每一方向的反应条件并不相同。这是噬菌体生命过程中的一个重要特征，因为它提供一种方式来保证整合的过程并不会被切除过程所立即逆转，反之亦然。

整合与切除反应所需不同的反应位点对，原因是参与此两个过程的蛋白质不同。整合反应由 λ 噬菌体基因 *int* 上的产物整合酶(integrase，Int)催化。Int 只能催化 *BOB′* 和 *POP′* 之间的重组。因此在只有整合酶存在时，上述反应是不可逆的。Int 是一种 DNA 结合蛋白，对 *POP′* 序列有强的亲和力。整合反应还需要一种由大肠杆菌编码的一种细菌蛋白，称为整合宿主因子 IHF(integration host factor，IHF)，IHF 是一个含两个不同亚单位的 20 kDa 的蛋白，此两亚单位分别为 *himA* 和 *himD* 基因所编码。IHF 也与 *att* 位点结合。Int 和 IHF 可以在体外进行位点特异性重组。在没有任何 DNA 合成的情况下，它涉及到 DNA 的精确的割裂和重新接合。可以通过在任意一侧对 *att* 位点的缺失来研究它们的作用，结果显示，*attP* 要比 *attB* 大得多。*attP* 的功能要求一段 240 bp 的 DNA，但 *attB* 的功能可以被一个从 −11～+11 位点的 23 bp DNA 片段执行，并且在核心序列的两侧只有 4 bp。这种大小的差异表明，*attP* 和 *attB* 在重组中所起的作用不同，与 *attB* 相区别的是，*attP* 提供附加的信息。

切除反应发生在原噬菌体两端的 *attL*(*BOP′*)和 *attR*(*POB′*)之间，产物为 λ 噬菌体环状 DNA 和细菌染色体 DNA。重组后 λ 噬菌体的 *att* 位点恢复为 *attP*(*POP′*)，细菌

DNA 的 *att* 位点为 *BOB′*。催化切除反应的除了 Int 和 IHF 外，还需要一种叫做切除酶(exicisionase，Xis)的蛋白参加，该酶由 λ 噬菌体的 *xis* 基因编码。整合和切除反应皆需 Int 和 IHF 蛋白。Xis 对控制反应方向起重要作用，它是切除反应所需要的，但却抑制整合反应。Xis 和 Int 结合形成复合体，该复合体具有与 *BOP′* 和 *POB′* 结合的能力，促进两者之间相互作用和重组，Xis 和 Int 不能催化 *BOB′* 和 *POP′* 之间的重组。故在 Xis 大量存在时，切除作用是不可逆的。

λ 噬菌体的整合和切除反应是受到严格的遗传学控制的，当 λ 噬菌体进入溶源状态时，发生整合反应所需的 Int 蛋白质合成，*xis* 基因失活，这样就保证了在溶源化的过程中 Int 蛋白起作用，无 Xis 蛋白存在。而当切除反应发生时，*xis* 和 *int* 基因转录产生 Xis 和 Int 蛋白，催化切除反应。位点特异性重组的复杂性大多产生于对反应的调控，使其在病毒进入溶源状态时发生整合反应，并且当原噬菌体进入裂解循环时，发生切除反应。通过控制 Int 和 Xis 蛋白的数量，适当反应就可以发生。

可以把整合和切除反应写成：

$$\underset{\text{噬菌体}}{POP'(attP)} + \underset{\text{细菌}}{BOB'(attB)} \underset{\text{Int、IHF、Xis}}{\overset{\text{Int、IHF}}{\rightleftharpoons}} \underset{\text{原噬菌体}}{BOP'(attL) + POB'(attR)}$$

λ 噬菌体的整合和切除反应见图 9－12。

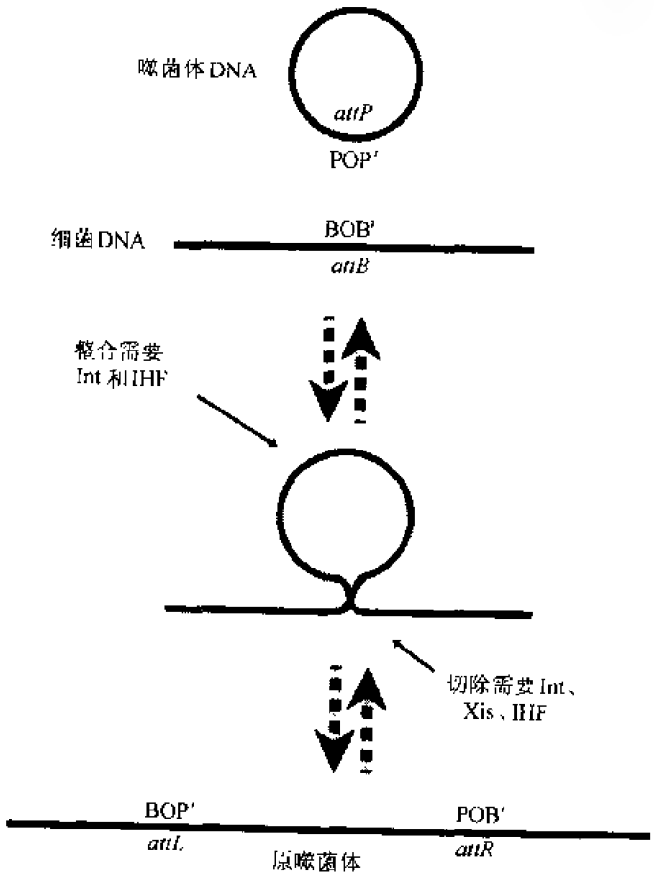


图 9－12 *attP* 和 *attB* 位点间的重组反应模式
通过在 *attP* 和 *attB* 间的交互重组，环状的噬菌体 DNA 转换为整合的原噬菌体，原噬菌体通过 *attL* 和 *attR* 间的交互重组而被切除。

二、λ 噬菌体 DNA 的整合机制

重组反应可以被“自杀底物”(suicide substrate)作用而停止在反应的各中间阶段。在此情况下核心序列含有断点，断点的存在会对重组起干扰作用。这使对已发生重组但尚未完成的分子进行鉴定成为可能，这些中间体的结构表明链交换所发生的顺序。λDNA 的整合反应涉及到 *attB* 和 *attP* 的核心序列中链的割裂与重接(图 9－13)。首先在 *attP* 和 *attB* 位点上产生同样的交错切口，形成了 5′－OH 和 3′－P 的末端。5′单链区全长 7 个碱基。同位素标记试验证明，两个核心区的断裂完全相同，连接过程不需要任何新 DNA 的合成。在整合反应中，互补的单链末端交互杂合，连接并完成整合过程。

在离体条件下，Int 和 IHF 可

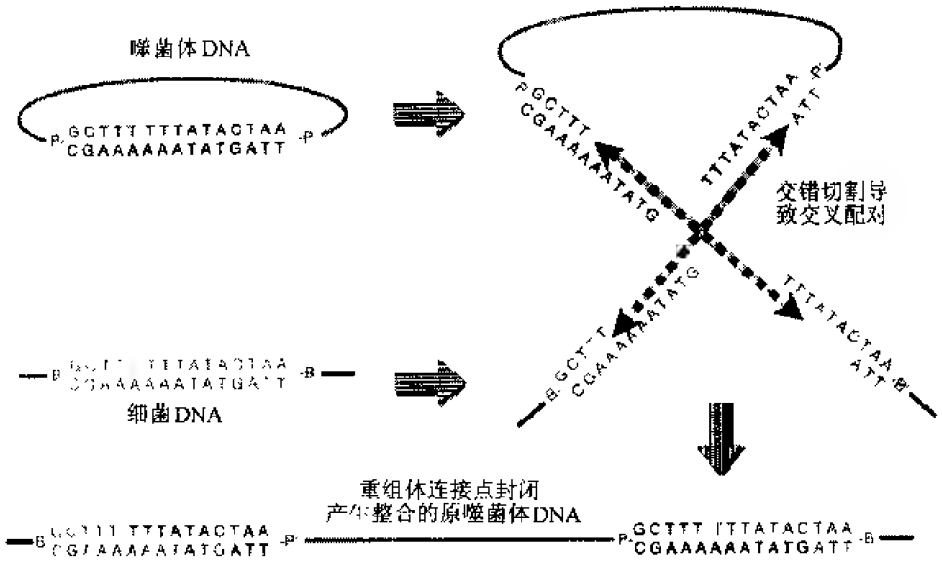


图 9-13　噬菌体整合过程的分子机制
在 *attP* 和 *attB* 共有的核心序列上交错切割，交错切割导致交互配对，重组体末端连接，产生整合的原噬菌体。

以催化 λDNA 和宿主 DNA 进行位点特异性重组，但需要在 *attP* 上有超螺旋，而 *attB* 则不需要。在体外，当反应发生在两个超螺旋 DNA 分子之间时，产物几乎保留所有的超螺旋。说明整个反应中没有可以自由旋转的中间体，断裂和重新结合反应与拓扑异构酶 I 催化的反应机制相类似，不同的是相互连接的两条链来自于不同的双螺旋，而不是来自于同一双螺旋（Int 确实具有相当低效率的使负的超螺旋松驰的拓扑异构酶 I 活性）。

在体外，重组反应需要大量的 Int 和 IHF 蛋白，Int 和 IHF 以协同的方式与 *attP* 相结合，并且它们对位点的亲和性会被超螺旋所增强。这种高比例表明，蛋白质很可能与维持支持重组过程的结构有关。

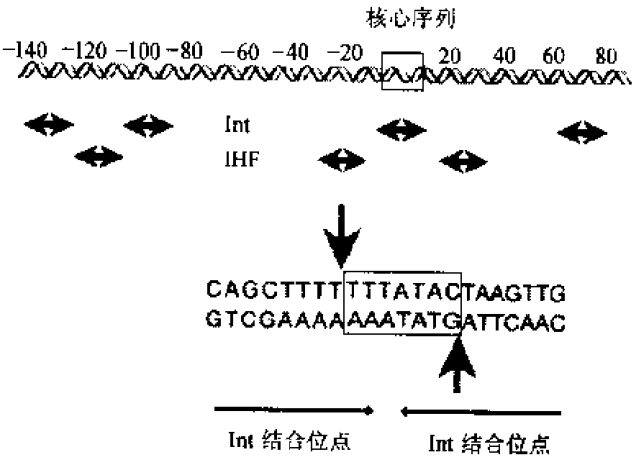


图 9-14　Int 和 IHF 在 *attP* 上的结合位点

参与位点特异性重组的蛋白质在 *att* 部位与 DNA 的一些特异位点相结合。Int 存在两种不同的结合方式，如图 9-14 所示，它与核心序列两个反向位点相结合，这种定位可使其切断两条相应的单链。这些位点具有共有序列（consensus sequence）。另一种位于 *attP* 臂上，这些位点没有共有序列。Int 蛋白上的不同结构域识别不同的序列。一个 N 末端的结构域识别 *attP* 臂上的位点，一个 C 末端的结构域识别 *attB* 和 *attP* 核心序列。这两个结构域大概同时与 DNA 结合，使 *attP* 臂上位点与核心区靠近。

IHF 与 *attP* 上一个大约 20 bp 的序列相结合，IHF 结合位点大概与 Int 的结合位点

相邻。Xis 与 *attP* 上两个位置接近的位点相结合，因此被保护的区域可扩展至 30～40 bp。Int、Xis 和 IHF 实际上覆盖了 *attP*。Xis 的结合可改变 DNA 的结构，使其对整合反应呈惰性。

当中心序列与 Int 相结合时，所有的接触点皆位于 DNA 分子的一侧，两个切点被暴露于大沟，IHF 的结合位点位于同一表面。如果 Int 结合位点与两侧的 IHF 结合位点之间的距离被改变使其不再是 DNA 螺旋的整数倍时，整合受到阻碍。

当 Int 和 IHF 与 *attP* 相结合时，它们形成一个复合物，该复合物称为整合体(intasome)。在整合体中所有的结合位点都被集中到此蛋白寡聚体的表面，此整合体的形成需要 *attP* 超螺旋。

attB 的结合位点是位于核心序列的两个 Int 位点，但 Int 并不与以游离形式存在于 DNA 上的 *attB* 直接结合，整合体是"捕获"*attB* 的中间体，大概是作为整合体中一部分的 Int 分子与 *attB* 的核心序列上的结合位点相结合。

根据这一模型，对 *attP* 和 *attB* 的初始识别并不直接依赖于 DNA 序列的同源性，但却依赖于 Int 蛋白能识别两者的 *att* 序列的能力，依据整合体的结构，两个 *att* 位点按预定的方向被带到一起。接着要发生链的交换反应，此时序列同源性则显示了其重要性。

第六节　异常重组[78～87]

异常重组(illegitimate recombination)，发生在彼此同源性很小或没有同源性的 DNA 序列之间。这种重组过程可发生在 DNA 很多不同的位点，它们可能是最原始的重组类型，不需要对特异性序列进行识别的复杂系统或对 DNA 同源序列进行识别的机制。这些重组过程和癌症发生、遗传性疾病和基因组进化有关。

异常重组按其机制主要分为两类：末端连接(end-joining)和链滑动(strand-slippage)。末端连接是指断裂的 DNA 末端彼此相连。链滑动是指 DNA 复制时，由一个模板跳跃到另一个模板所引起的重组。异常重组能够产生许多不同的结果，例如移码、缺失、倒位、融合和 DNA 扩增。

许多 DNA 序列都可发生上述两类的异常重组，这直接威胁着基因组的完整性，但同时也是进化的重要途径。

一、末端连接

真核细胞末端连接是一个高效率的反应。在 40 年代初首次被 Barbara McClintock发现。她观察到如果玉米细胞遗传一个含断裂末端的染色单体，这一染色单体能够进行复制，两个断端可以融合产生一个双着丝粒的染色单体。该染色单体在减数分裂时断裂并产生断端，断端再次融合，又产生一个新的双着丝粒的染色单体，在下次的减数分裂时又将断裂，这一过程称为断裂-融合-桥(breakage-fusion-bridge，BFB)循环。这一循环将持续许多轮，直到端粒加到断端上为止(图 9－15)。已表明 BFB 循环是与染色体的某些区域以头对头的方式连接有关，产生上兆个碱基的回文区(图 9－15)。

David Roth 等用猴细胞模型系统、Petra Pfeiffer 用非洲爪蟾卵提取物的类似系统对

末端连接的机制进行了广泛的研究。断裂末端可以通过断端直接对接相连[图 9－6(1)]。然而，如果断端不能直接相连(如两末端的极性相同时)，它们可以通过碱基配对，然后末端合成修复[图 9－16(2)]。断端连接后产生新的连接体。该过程和末端 DNA 的序列无关，也不要求两断端之间有任何同源性。如果有微小的同源区配对，反应发生在短的正向重复处，典型的为 2～5bp 长。此外，大约 10% 的情况下，在断端之间发现有随机的核苷酸。

在大肠杆菌，末端连接反应要求有一小段能进行配对的 DNA 同源区。这是因为大肠杆菌不能直接连接单链 DNA 末端。当线形质粒分子进入细胞时，能够用线形化时所产生的末端重新环化。如果不发生环化，则可能发生两种反应。末端能够与有一小段同源碱基的内部序列结合，或与有一小段同源区的两个适宜的内部序列结合。目前认为经核酸酶消化产生单链 DNA 后，可产生短的同源区。

二、链 滑 动

1966 年 George Streisinger 提出，通过新合成的 DNA 链与其模板的错配可导致移码突变。同样机制在较大的范围内发生时，可以产生 DNA 的缺失和扩增，其长度是由新、旧模板之间的距离而定(见图 9－17)。因为该机制涉及到新复制的链同模板链的配对，这些过程可能通过 Watson-Crick 碱基配对而得到促进，因此它常发生在短的正向重复序列。

某些正向重复序列似乎是链滑动发生的热点，Jeffrey Miller 等实验室的研究已经在大肠杆菌的 *LacI* 基因鉴定了一个突变的热点，所有自发性突变的 2/3 由它引起，它由 5′CTGG3′和它的互补链 5′CCAG3′组成，以串联的形式重复 3 次。这种四核苷酸序列拷贝数的增加和减少已被发现，表明环出(looping out)可以发生在模板链和新合成的链。在人类染色体上类似的

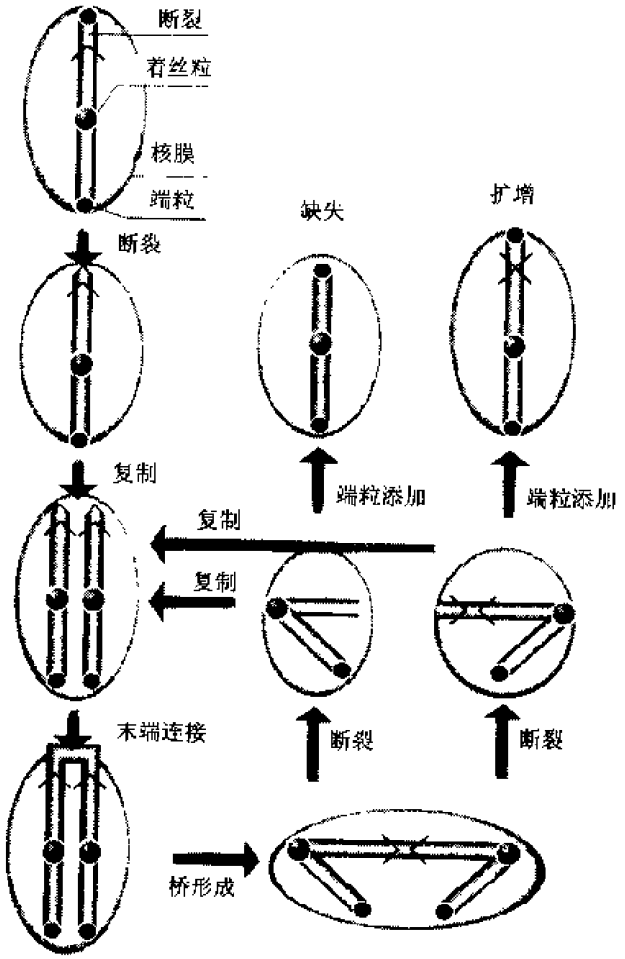


图 9－15 断裂-融合-桥循环

断裂-融合-桥循环过程中 DNA 序列能被扩增，形成一个长的 DNA 回文结构。图示一个染色体的遗传标志(用箭头示)和端粒之间发生断裂。在减数分裂时，染色体复制产生两个含断端的染色单体，两个染色单体通过末端连接融合在一起。在减数分裂的后期，含双着丝粒的染色单体被牵拉，移向两极。桥断裂，两个产物进入两个子细胞核。断裂可能是非对称的并产生一个序列缺失，一个序列扩增的染色体。扩增的染色体在末端含一个长的回文结构，它是在末端连接时形成的。缺失和扩增的染色体仍然含有断端，并能再次进入 BFB 循环，循环直到端粒被加到断端时终止。

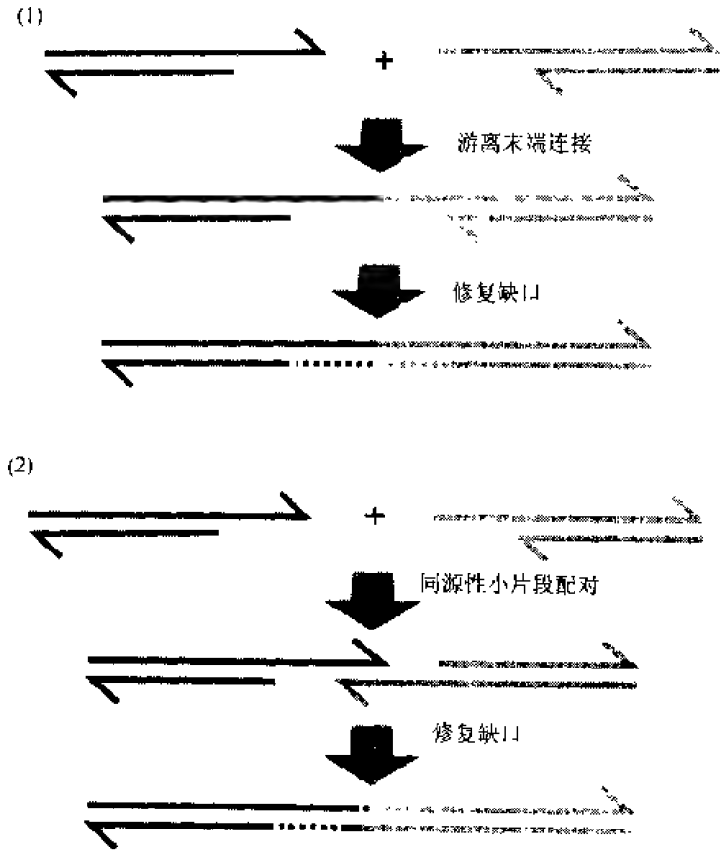


图 9-16　末端连接反应

在哺乳动物细胞，末端连接反应能够通过两种途径进行。图(1)任意两个末端彼此相连。图(2)在连接前氢键配对。上述两种情况有少量的 DNA 合成。

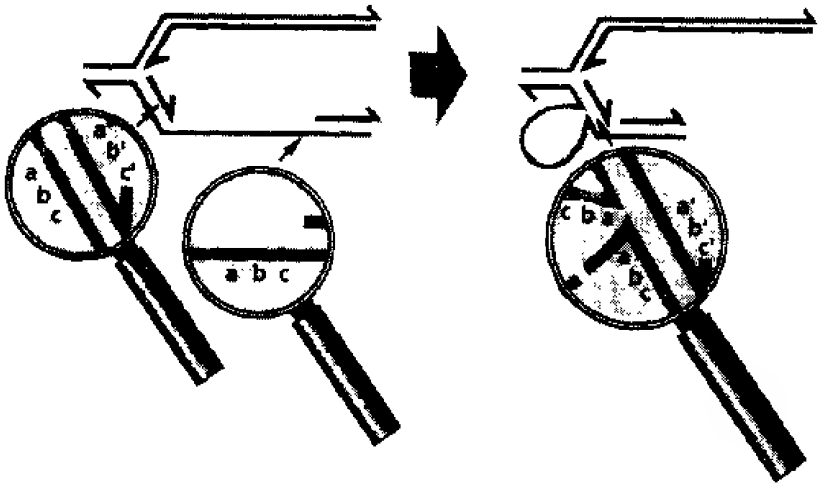


图 9-17　通过链滑动引起缺失

在 DNA 复制过程中，新合成的 DNA 链有时与它的模板解离，并与下游的互补序列配对，互补序列为 a、a′，b、b′和 c、c′。这可导致 DNA 序列缺失。

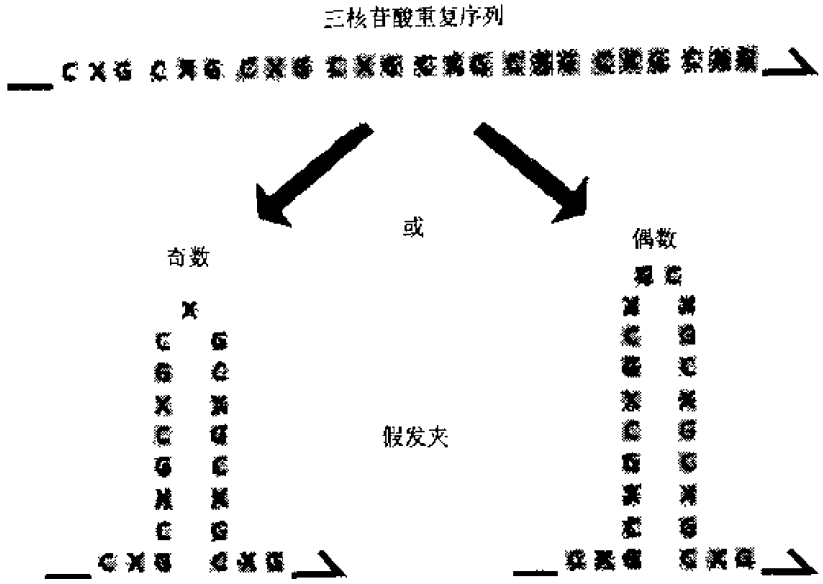


图 9-18　三核苷酸重复序列形成的假发夹

三核苷酸重复序列能形成一个或多个特殊的二级结构，由于链内的 C 和 G 碱基之间的配对而稳定。图示奇数或偶数重复序列形成的结构，这样的假发夹结构在体外寡聚核苷酸的研究中已被发现，并且在体内的资料也有报道。

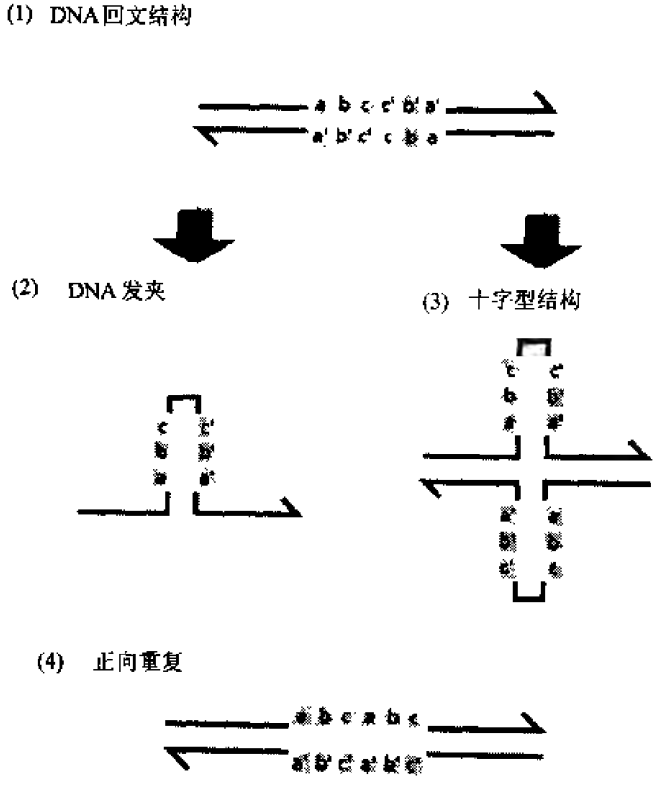


图 9-19　回文结构、发夹结构、十字形结构和正向重复

碱基 a、b 和 c 分别与碱基 a′、b′和 c′互补，(1)所示为一线性的回文结构表示形式。(2)所示为回文结构的一条链发生链内的碱基配对，产生一个发夹结构。(3)所示为回文结构的两条链发生链内碱基配对，产生一个十字型结构。(4)所示为一个正向重复序列没有二级结构形成。

缺失和扩增的突变和遗传性疾病发生有关。这些突变发生在3个或4个核苷酸的重复序列中，称为动态突变(dynamic mutations)，5′CCG3′序列在正常人为6～52个拷贝，在脆性X染色体综合征患者体内可多于200个拷贝。在亨廷顿氏病(Huntington's disease，HD)，5′CAG 3′序列由6～33个拷贝增加到35～121个拷贝。目前认为单链的5′CAG3′序列有形成异常二级结构的倾向。这可能是一种假发夹，并通过G/C碱基配对稳定(如图9-18)。此外，这些序列难于通过DNA聚合酶进行复制，这可能由于停顿过长，从而促进链滑动。在动态突变中，尚未证明链滑动的作用。在另一种模型中，由于同源重组引起重复序列数目的增加和减少也是可能的。

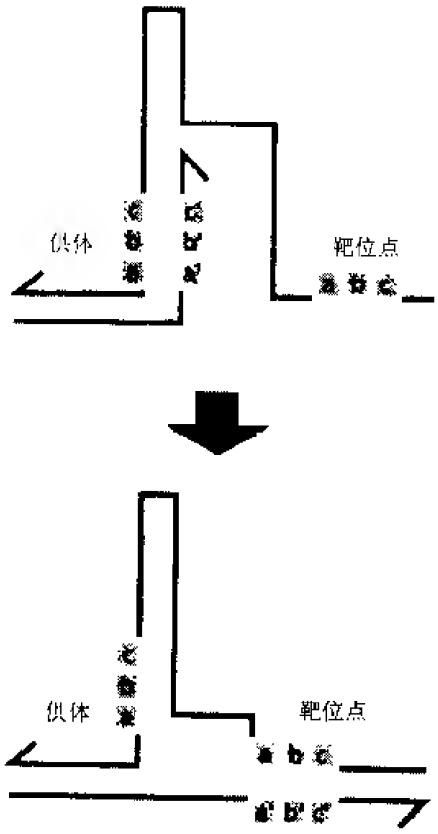


图9-20 回文结构引发复制时的链滑动
当DNA复制进入到链内碱基配对区时，有可能产生一个停顿。新合成链的3′端从模板解离并与下游的另一个互补序列配对。这就是所谓供体重复序列(在回文结构的内部)和靶重复序列(在回文结构的外部)。这些正向重复(由a、b和c代表)可以小到只有3个核苷酸长。

末端连接和链滑动反应能导致在短的正向重复序列处形成重组连接点。同样，如果在链滑动的过程中有非Watson-Crick氢键的相互作用的可能性，通过末端连接和链滑动这两种机制有可能在非同源位点形成连接。因此通过产物的结构区别这两种机制是不可能的。末端连接反应涉及到通过双链切割产生的末端之间的相互作用，而链滑动涉及到的只是一个单链末端与完整的链之间的相互作用。在DNA被重新引入到细胞内之前，在体外使双链断裂，是一个很有用的研究末端连接的实验技术。然而研究链滑动则没有这样简单的方法。因此，对它参与大规模重排的许多争论的证据都是间接的。

三、二级结构促进异常重组

有充分的研究证明，能引发异常重组的DNA序列是反向重复序列和长的回文序列。回文序列是没有间隔DNA的反向重复序列。这些序列有可能以单链发夹结构的形式存在，或以双链的形式形成十字型结构(如图9-19)，这些结构可以通过链滑动引发异常重组。通过与其末端相邻的、短的正向重复序列之间的重组，回文结构被去除。最好是一个重复序列位于回文结构内部，而另一个重复序列又刚好位于回文结构外部，这就构成了一个模式，复制进行到回文结构时只进行到一定的距离之后则停滞。然后新合成的链从模板上解离并与附近的任何互补序列配对。这种典型序列常见于回文序列的下游。我们可以根据这种不对称性来确定缺失反应中的供体和靶重复序列。供体重复序列位于回文序列内部，而靶重复序列位于回文序列的外部(如图9-20)。这种不对称性通常用于检测链滑动更常见于复制叉的前导链还是后随链，这可以通过确定一段含有回文结构和适宜位点的DNA的倒转效应来完成(如图9-21)。资料表

明，缺失常发生在后随链。

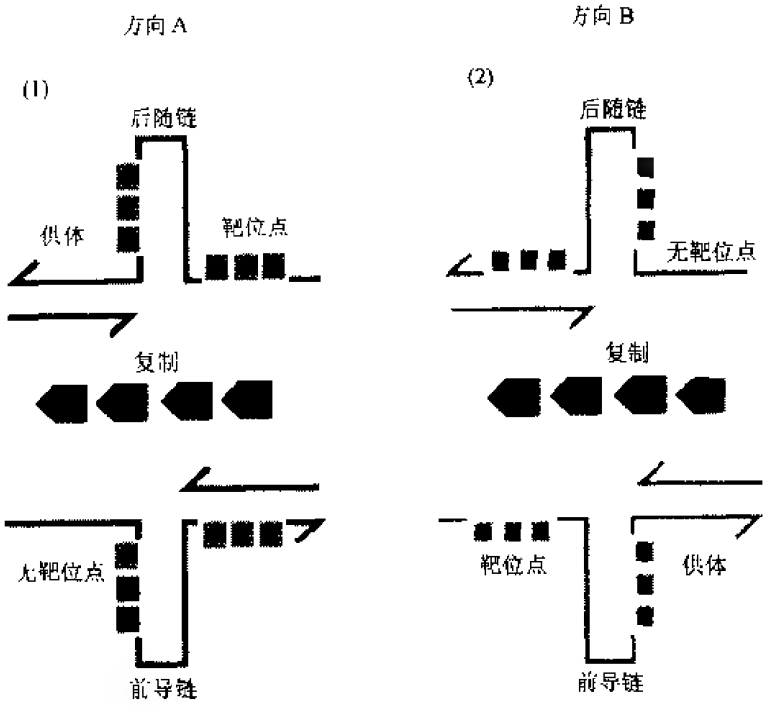


图 9－21　链滑动常发生在后随链

当一个新复制的 DNA 链从一个方向进入到一个回文序列时，如果在回文序列的下游有一个适宜的靶重复序列，链滑动可能发生。一个适宜的靶重复序列是恰好位于回文结构之外的一个序列，与回文结构内的序列有几个碱基互补。图(1)适宜的靶位点位于后随链复制的下游，而图(2)中，同样适宜的靶位点位于前导链复制的下游。资料表明，如果链滑动发生，优先发生在后随链。(1)的缺失频率应当较高。

在重复序列之间带有非重复序列的反向重复序列也可发生缺失，但与较典型的回文序列相比，其频率较低。这些异常重组反应引起转座元件切除(例如 Tn*10* 和 *P*)，其作用机制可能与此相似，在大肠杆菌中很长的、典型的回文序列(总长度大于 150 bp)，根本不能被复制。这是由于 Sbc CD(一种依赖于 ATP 的核酸酶)的作用下切除了 DNA 的发夹结构。

(张玉静)

参 考 文 献

[1] Lewin B. Genes, New York: Oxford University Press, 1997. 989
[2] David R F. Genetic recombination. Blackwell Science Ltd, 1996
[3] Snyder L et al. Moleculat Genetics of Bacteria. American Society for Mrcrobiology. Washington D C, 1997
[4] 廉克中. 基因及其表达. 北京：科学出版社，1996
[5] Strachan Tom et al. BIOS Scientific Publishers Limited, 1996
[6] Low K B. A Collection of Review Articles on Homologous Recombination. Academic Press, San Diego, 1988
[7] Myers R S et al. Ann Rev Genet, 1994. 28: 49
[8] Petes T D. The Molecular and Cellular Biology of Yeast Saccharomyces. Cold Spring Harbour Press, 1991. 1: 407
[9] Stahl F W. Genetic Recombination, Thinking About it in Phage and Fungi. Freeman, USA, 1979

[10]Whitehouse H L K. Genetic Recombination, Understanding the Mechanisms. Wiley, New York, 1982
[11] Berg D E et al. Mobile DNA. American Society for Microbiology, Washington, 1989
[12] Griffiths A J et al. An Introduction to Genetic Analysis(5th edn). Freeman, New York, 1993
[13] Kulcherlapati R et al. Genetic Recombination. American Society for Microbiology Washington, 1988
[14] Moens P B et al. Meiosis. London: Academic Press Orlando, 1987
[15] Stahl F W. Ann Rev Genet, 1979. 13:7
[16] Simon M et al. Science, 1980. 209:1370
[17] Dressler D et al. Ann Rev Biochem, 1982. 51:727
[18] Cox M M et al. Ann Rev Biochem, 1987. 56:229
[19] Glasgow A C et al. Mobile DNA. Washington D. C.: American Society for Microbiology, 1989
[20] Carpenter A T. Proc Nat Aacad Sci USA, 1975. 72:3186
[21] Shibata T R et al. Proc Nat Aacad Sci USA, 1979. 76:1638
[22] Szostak J W et al. Cell, 1983. 33:25
[23] Padmore R et al. Cell, 1991. 66:1239
[24] Sun H et al. Cell, 1991. 64:1155
[25] Holliday R. Genet. Res, 1964. 5:282
[26] Jones M et al. Cell, 1987. 50:621
[27] Paquette N et al. Molec. Gen. Genet, 1978. 163:313
[28] Holliday R. Genet Res, 1964. 5:282
[29] Kowalczykowski S. Microbiol. Rev, 1994. 58:401
[30] Lloyd R G et al. Cellular and Molecular Biology, 2nd ed. Washington, D. C.: ASM Press, 1996
[31] Meselson M S et al. Proc Natl. Acad Sci. USA. 1975. 2:358
[32] Lloyd R G. J Bacteriol, 1991. 173:5414
[33] Lloyd R G. J Bacteriol, 1985. 164:836
[34] Dixon D A et al. Cell, 1993. 73:87
[35] Miesel L et al. J Bacteriol, 1996. 178:3146
[36] Parsons C A et al. Proc Natl Acad Sci USA, 1992. 89:5452
[37] Rao B J et al. Proc Natl Acad Sci USA, 1991. 88:2984
[38] Sawitzeke J A. Genetics, 1992. 130:7
[39] Stahl F W et al. Genetics, 1980. 94:235
[40] Szostak J W et al. Cell, 1983. 33:25
[41] Taylor A F et al. Proc Natl Acad Sci USA, 1992. 89:5226
[42] Taylor A F. Genetic Recombination, Washington: American Society for Microbiology, 1988. pp. 231～263
[43] Thaler D S et al. Genetics, 1987. 116:501
[44] West S C. Proc Natl Acad Sci USA, 1981. 78:6149
[45] Cox M M et al. Proc Natl Acad Sci USA, 1981. 78:6018
[46] Hagemann A T et al. Genetics, 1991. 129:611
[47] Kahn R et al. Proc Natl Acad Sci USA, 1981. 78:4786
[48] Kobayashi I et al. Proc Natl Acad Sci USA, 1982. 79:5981
[49] Kobayashi I et al. Cold Spring Harbour Symp. Quant. Biol. 1985. 49:497
[50] Kobayashi I et al. Genetics, 1983. 104:549
[51] Potter H et al. Proc Natl Acad Sci USA, 1977. 74:4168
[52] Rao J J. Proc Natl Acad Sci USA, 1993. 90:6646
[53] Siddiqui I. Genetics, 1991. 128:7
[54] Stahl F W. Mole Gen Genet, 1975. 140:29
[55] Stahl F W. J Mole Biol, 1985. 181:199
[56] Stahl F W et al. Genetics, 1974. 77:395
[57] Cozzarelli N R et al. DNA Topology and its Biological Effects, New York: Cold Spring Harbor Press, 1990. PP. 139～184
[58] Droge P et al. Methods in Enzymology, 1992. 212:120
[59] Drlica K. Microbiol Rev, 1984. 48:273

[60] Wang J. Ann Rev Biochem, 1985. 54:665
[61] Wang T C. Ann Rev Biochem, 1996. 65:635
[62] Cox M M. Genetic Recombination. Washington: American Society for Microbiology, 1988. 429～443
[63] Hall R M et al. Molec Microbiol, 1995. 15:593
[64] Hatfull G F Genetic Recombination. Washington: American Society for Microbiology, 1988. 357～396
[65] Landy A. Ann Rev Biochem, 1989. 58:913
[66] Stark W M, Boocock M R et al. Trends Genet, 1989. 5:304
[67] Stark W M. Trends Genet, 1992. 8:320
[68] Stark W M et al. Mobile Genetic Elements: Frontiers in Molecular Biology. Oxford: Oxford University Press, 1995. 101～129
[69] Benjamin H W et al. Cell, 1985. 40:147
[70] Cox M M. Genetic Recombination. Washington: American Society for Microbiology, 1988. 429～443
[71] Nash H A et al. J Mol Biol, 1983. 170:19
[72] Reed R R. Cell, 1981. 25:713
[73] Spengler S J et al. DNA structure Cell, 1985. 42:325
[74] Stark W M et al. Cell, 1989. 58, 779～790.
[75] Nash H A. Cellular and Molecular Biology, 2nd ed. Washington, D. C.: ASM Press, 1996
[76] Weisberg R et al. Site-specific recombination in phage lambda. New York: Cold spring Harbor Laboratory, 1983
[77] Thompson J F et al. Mobile DNA. Washington D. C.: American Society for mocrobiology, 1989
[78] Allgood N D et al. Genetic Recombination. Washington D. C.: American Society for Microbiology, 1988. 309～330
[79] Champoux J J et al. Genetic Recombination. Washington: American Society for Microbiology, 1988. 655～667
[80] Ehrlich S D. Mobile DNA. Washington: American Society for Microbiology, 1989. 799～832
[81] Ehrlich S D et al. Gene, 1993. 135:161
[82] Meuth M. Mobile DNA. Washington: American Society for Microbiology, 1989. 833～860
[83] Leach D R F. Genetic Engineering, Principles and Methods. London: Plenum Publishing, 1995
[84] Leach D R F. BioEssays, 1994. 16:893
[85] Roth D. Genetic Recombination. Washington: American Society for Microbiology, 1988. 621～654
[86] Darlow J M. Genetics, 1995. 141:32
[87] Trinh T Q et al. Nature, 1991. 352:544

第十章 转 座

第一节 概 述[1-8]

通常认为基因组是相对稳定的，变化只是在漫长的进化过程中发生，我们习惯的做法是通过遗传图谱的构建来鉴定已知基因的基因座，这意味着基因组中的序列通常处于恒定的位置。同源性重组允许同源染色体之间进行遗传物质交换，产生等位基因新的组合，除此之外，不对遗传物质进行重新组织。

物种分化之后其相互之间的关联性（例如：人和猿之间）表明遗传结构是相对稳定的。原核和真核细胞产生的时间上的差异表明它们在时间上的进化尺度也许是不同的，但即使对原核生物来说，基因组总的结构变化也是相对较慢的。例如，不同的细菌，如大肠杆菌和鼠伤寒沙门氏菌的遗传图谱相似。

同总的稳定性相反，由于多态变异体的存在，在分子水平上，每个个体的基因组之间具有明显的差异性。我们观察到，小卫星 DNA 的长度由于重组而发生变化，致使每个个体的基因组皆不相同。另一主要的变化，是由转座元件（transposable elements）或转座子（transposons）所引起的，它们是在基因组中可移动的独立的 DNA 序列，也就是能够转移到基因组内的其他位置上去。转座子的标志是它不以独立的形式存在（如噬菌体或质粒 DNA），而在基因组内由一个部位直接转移到另一个部位。一个转座子由基因组的一个位置移到另一个位置的过程称为转座。促进转座的酶称转座酶，转座子常常编码它自己的转座酶，并且每次移动时携带转座必需的基因一起在基因组内跃迁，所以转座子又称跳跃基因（jumping gene）。转座以很低的频率发生，而且转座子的插入是随机的，不依赖于转座子和靶位点之间任何序列的同源性（有时有程度不同的倾向性）。转座子有时插入到一个结构基因或基因调节序列内并引起基因表达的改变。

转座元件首先是由 Barbara McClintock 于 40 年代在玉米的遗传学研究中发现的，她称之为控制元件（controlling element），因为它不仅能够在基因组内转移，而且也能改变基因的活性并引起功能改变。但在当时，这些发现未引起足够重视。直到 1968 年，分子水平的研究证实了在大肠杆菌中转座元件的存在，这才引起科学家的重视并进行了深入细致的研究。现在认为，转座子存在于地球上所有生物体内，包括人类。

有时转座子也引起基因组序列的重排。研究转座子的意义不仅仅在于其涉及到 DNA 的操作机制，它们的移动性也和进化有关，在基因组内，它们也许是突变的主要来源。

基因组的进化以获得新的序列和对原有序列进行重排的方式进行。新的序列的突然引入是载体在基因组之间转移信息的结果。染色体外的成分通过介导一定长度（通常很短）的遗传物质的传递来水平转移信息，细菌通过接合转移质粒，而噬菌体通过感染传播。质粒和噬菌体在转移它们自己的复制子的同时偶尔也会转移宿主的基因，某些细菌通过转化来直接交换 DNA。在真核生物中，某些病毒（例如反转录病毒）在感染周期中能传递

遗传信息。

重排可能产生新的序列，并且通过置原有序列于新的调节状态下而改变其功能。重排是基因组内部发生的过程。它们是细胞系统由于同源性重组和修复而进行重组或转座子转座的结果。

1）交互重组发生在真核细胞内同源染色体相关部位之间；偶尔发生的非交叉重组引起基因座的复制或重排。在一个基因组内序列的复制是新序列的主要来源。该序列的一个拷贝可以保留它的原有功能而另一个可以产生一种新的功能。在一个染色体内，这种重新组合显然是和遗传重组有关的通常反应机制的相关产物。大量的重排也见于非同源染色体之间的易位，但还不清楚其产生机制。

2）在原核和真核基因组中，转座子的移动性提供了产生变化的强大动力，和基因组重新构建有关的许多其他过程不同，转座不依赖于供体和受体位点处序列之间的任何关系。转座子被限定于移动其本身（有时伴随附加序列）到同一基因组内的其他部位。因此，对于那些能把序列由一个基因组转移到另一个基因组的载体而言，它们是一种内部的载体类似物。

第二节　转座子的分类[31~40]

转座子分为两个基本类型，第一部分所描述的转座子是以编码蛋白质的 DNA 序列的形式存在，这些蛋白能直接操作 DNA，从而使转座子能够在基因组内繁殖其自身。下一部分叙述的转座子与反转录病毒相关，并且它们可移动的特性来自于它们能由 RNA 转录物产生 DNA 拷贝，这种 DNA 拷贝然后被整合到基因组的新位点。

在原核和真核细胞内都发现了通过 DNA 移动的转座子的存在。每个细菌的转座子皆携带着编码其自身转座所需的酶活性的基因，虽然它也可能需要其所在的基因组的辅助功能（如 DNA 聚合酶或旋转酶）。类似的系统也存在于真核细胞，虽然涉及到转座的酶的功能还尚未被很好地鉴定。

一个基因组可能含有功能性的和非功能性的（缺陷性的）转座成分；在真核基因组中，这些成分的大部分常常是缺陷性的，并且已经丧失了独立转座的能力，虽然它们仍然可以作为转座的底物而被功能性转座子产生的酶所识别。

转座元件可直接或间接地促进基因组的重排。转座过程本身可引起缺失、倒位或导致宿主的部分序列迁移到新的位置。由于转座子的“可移动的同源区”（portable regions of homology）的功能，使其可作为细胞重组系统的底物；在不同位置上的一个转座子的两个拷贝（甚至在不同的染色体上）可以提供交互重组的位点。这样的交换可引起缺失、插入、倒位或易位。

转座子的间歇的活动性似乎为自然选择提供了一个不确定的靶部位。这种考虑促成了一种看法认为（至少某些）转座元件既不赋予表型以优点也不赋予以缺点。它们构成“自私 DNA”（selfish DNA），只关心自身的繁殖。的确，如果考虑转座为一种不同于细胞其他重组系统的过程，我们还是接受了这种观点：转座子是位子基因组内的一个独立的实体。

转座子与基因组的这种关系相当于寄生虫与宿主的关系。转座子的扩增大概通过产

生有害的结果而得到平衡，例如转座过程可能火活一个必需的基因，或者过多转座子的存在可变为细胞的一种负担。然而我们也必须记住，任何转座过程皆可提供一个选择优势（例如，基因重排），这将导致携带活性转座子的基因组易于存活。

一、插入序列

最早被鉴定的转座元件是细菌操纵子中的自主插入序列。这种插入阻止了插入部位基因的转录和/或翻译。到目前为止，已经对许多不同类型的转座元件进行了鉴定。

最简单的转座子被称为插入序列 IS(insertion sequences，IS)。每个类型皆被给予一个 IS 前缀，其后是一个数字代表其类型。（最早的类型被编码为 IS*1*～*4*；后来编码的类型后面的数字表明它们分离的历史，但和目前所分离的转座子的总数无关）。

IS 元件是细菌染色体和质粒的正常组成部分，一个大肠杆菌标准株很可能含有几个拷贝（一般小于 10 个）的最常见的任意一种 IS 元件。可利用双冒号来描述对一个特殊位点的插入，例如，λ::IS*1* 指的是一个 IS*1* 元件插入到 λ 噬菌体中。

IS 元件是独立的结构单位，每个元件只编码为自身转座所需要的蛋白质。每种 IS 元件在序列上皆是不同的，但在结构上具有一些共同的特征。一个转座子在靶部位插入前后的结构见图 10－1。

IS 元件的末端为短的反向末端重复(inverted terminal repeats)序列，通常这两个拷贝密切相关，但并不完全一样。如图 10－1 所示。

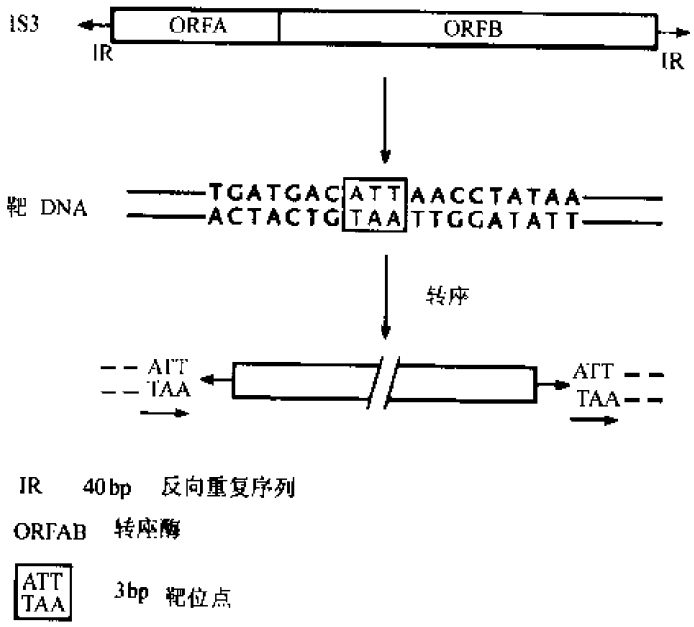


图 10－1 IS3 的结构
插入序列 IS3 的结构，箭头表示两端的反向重复序列。在靶部位产生正向重复序列。

当一个 IS 元件转座时，插入部位的一段宿主 DNA 序列被复制。通过插入前后靶部位序列的比较揭示了复制的性质。图 10－1 表明，在插入部位，IS DNA 总是和很短的正向重复序列(direct repeats)相连。但在插入之前靶部位只有这两个重复序列中的一个。在此图中靶部位序列为 $\frac{ATT}{TAA}$，转座后在转座子的两侧各存在一个此序列的拷贝。

通过一个转座子进行的转座，在每个转座过程中，正向重复序列皆可不同，但对任何特定的 IS 序列来说，此重复序列的长度是恒定的（这与转座的机制相关），其最常见的长度为 3～9 bp。

因此，IS元件具有一个典型的结构，它的末端为反向重复序列，而与其相连的宿主DNA的末端为短的正向重复序列。当在一段DNA序列中见到这种结构类型时，可被用来鉴别转座子。

多数IS元件可插入到宿主DNA的许多不同部位，但一些IS元件对某些特殊热点显示出（不同程度的）倾向性。

反向重复序列标明了转座子的末端。所有类型转座子的转座皆需末端识别。阻止转座的"顺式作用"（*cis*-acting）突变位于末端，末端被与转座有关的蛋白所识别，该蛋白称为转座酶（transposase）。

除IS*1*之外，所有的IS元件皆含有一个单一的长编码区，其起始于一端的反向重复序列之内，而刚好终止于另一端的反向重复序列之前或之内，它编码转座酶。IS*1*的结构较为复杂，有两个单独的可读框；在翻译过程中，要通过一个框架的迁移而使两个可读框皆被利用而产生一个转座酶。

不同转座元件的转座频率不同，总的转座频率为每种元件每代10^{-3}～10^{-4}次。单一靶部位的插入频率与自发性突变率水平相当（每代10^{-5}～10^{-7}次）。回复（通过IS元件的确切切除）常常很少发生，其频率为每代10^{-4}～10^{-10}次，这大约是插入的频率的千分之一。

二、复合型转座子

一些转座子除了具有与转座有关的功能外也携带药物抗性（或其他）标志，这些转座子被命名为Tn，并在后边带有一个数字。一类较大的转座子被称为复合型转座子（composite transposons），它的中心区携带药物抗性标志，两侧为由IS元件组成的臂。

臂可以正向或（更常见的）反向方式排列，因此，一个带着正向重复臂的复合型转座子具有如下结构：|臂L〉中心区|臂R〉。如果臂是反向重复的，则其结构为|臂L〉中心区〈臂R。箭头表示臂的方向，L和R则代表转座子遗传图谱的任意一个由左到右的方向。一个复合型转座子的结构见图10-2，在此图中还

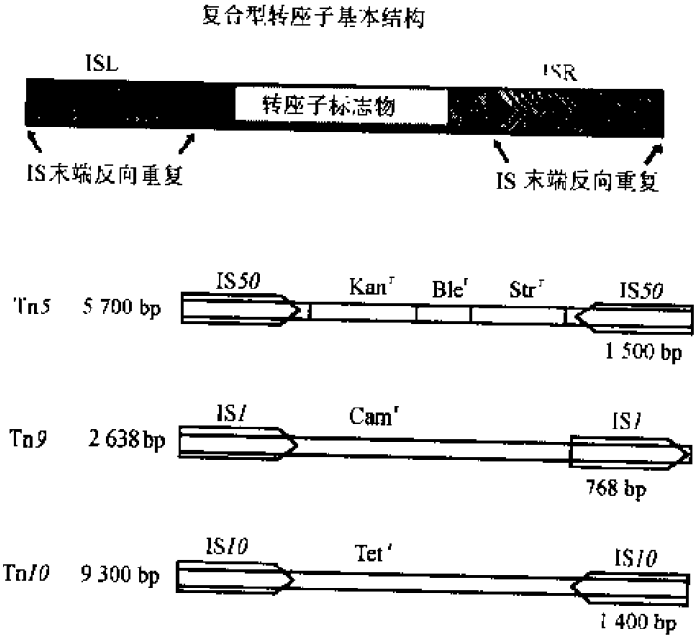


图10-2 复合型转座子的结构

某些复合型转座子的结构：复合型转座子有一个与转座无关的中心区（如药物抗性），两侧与IS元件相连Tn*5*、Tn*9*中含常见的基因为卡那霉素抗性基因*Kan*^{r}，氯霉素抗性基因*Cam*^{r}。活性转座酶的基因位于两个IS元件之中的一个。IS元件的排列可以是相同方向或相反方向的。Kan^{r}卡那霉素（kanamycin）抗性、Cam^{r}氯霉素（chloramphenicol）抗性、Str^{r}链霉素（streptomycin）抗性、Tet^{r}四环素（tetracycline）抗性和Ble^{r}博来霉素（bleomycin）抗性。

总结了一些常见的复合型转座子的特性。

因为臂由 IS 组成，每个 IS 具有通常的反向重复末端结构，因此复合型转座子具有同样的短的反向重复末端。

在某些情况下，一个复合型转座子的 IS 结构单位是一样的，例如，Tn*9*（IS*1* 正向重复）和 Tn*903*（IS*903* 反向重复）。在其他一些情况下，IS 单位是密切相关的，但并不完全一致。因此，我们能够区别出 Tn*10* 或 Tn*5* 的 L 和 R IS 序列。

一个功能性的 IS 结构单位能转座它本身或整个转座子。当一个复合型转座子的 IS 结构单位相同时，例如在 Tn*9* 或 Tn*903*，也许任意一个 IS 皆能使转座子移动。当 IS 不同时，它们也许在功能上也是不同的，例如在 Tn*5* 或 Tn*10*，这样转座完全或主要依赖两

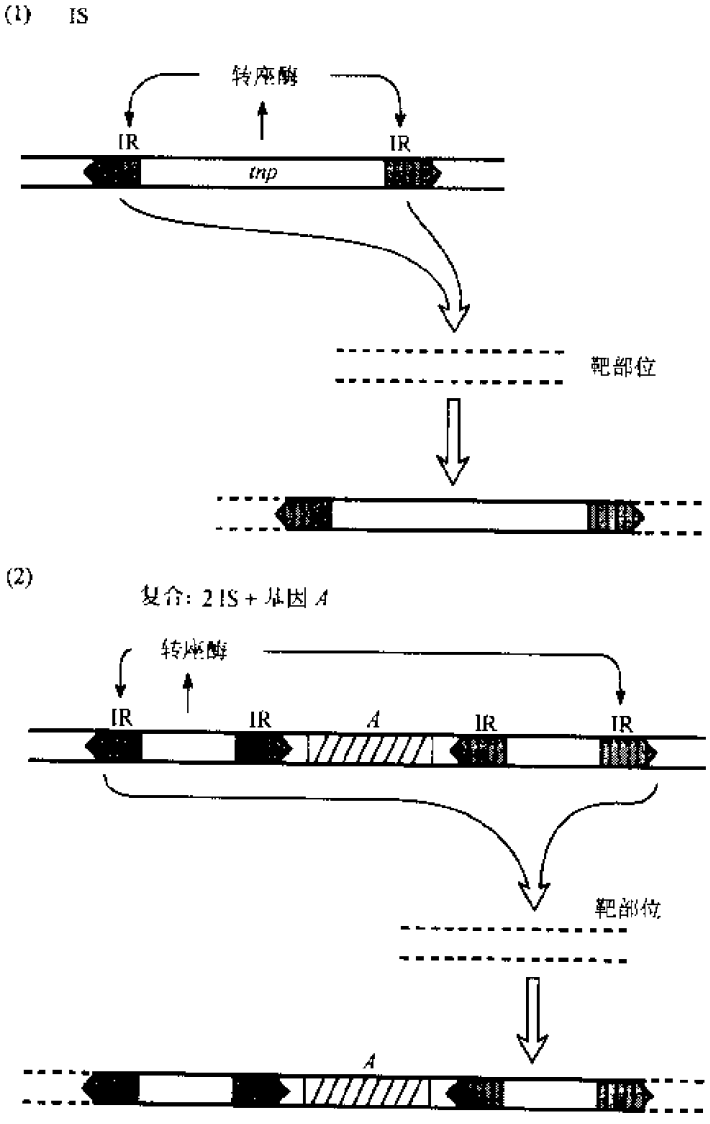


图 10-3 复合型转座子的转座

两个 IS 元件能够使任何位于它们之间的 DNA 序列转座。(1) 转座酶作用于 IS 元件的末端引起转座。(2)在 DNA 序列中，两个同样类型 IS 元件彼此靠近，转座酶在它们的外部末端发生作用，引起它们一起转座并携带着它们之间的 DNA 序列。

个 IS 中的一个。

我们假定当两个原来独立的 IS 结构单位与中心区连在一起时，即形成一个复合型转座子。当一个 IS 元件转座到邻近供体部位的一个受体部位时，这种情况即可出现。这两个相同的 IS 可以仍然保持相同或发生分化。一个单一的 IS 结构单位可使整个复合型元件座转的能力可解释为缺乏使两个 IS 结构单位皆保留活性的选择性压力。

是什么使复合型转座子而不是其 IS 结构单位发生转座呢？当两个 IS 皆有功能的情况下，该问题尤其显得急迫。在 Tn*9* 中，IS 结构单位为 IS*1*，假定每一个 IS 本身和复合型转 座子一样，皆具有活性，为什么转座子可作为一个整体得以维持而不是每一个插入序列自行其事呢？

事实上两个 IS 元件可与位于其中间的任何序列转座，包括其自身。如图 10－4 所示，如果 Tn*10* 位于一个环状的复制子上，它的两个 IS 结构单位可以被认为是与原来 Tn*10* 的 tet^r 基因相连或和环形结构的另一部分相连。这样一个转座过程可能涉及到原来的 Tn*10* 转座子或者产生一个带有另一个中间区域的内外颠倒的新的转座子。

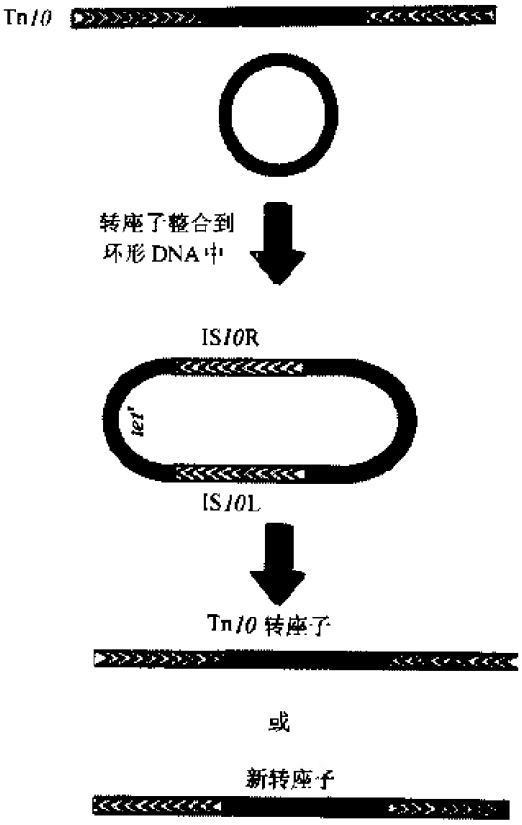


图 10－4　Tn*10* 的转座

两个 IS 元件产生一个复合型转座子，当 Tn*10* 为环形分子的一部分时，IS*10* 重复序列能使环的任意一侧转座。

注意原来的和内外颠倒的转座子皆有两个反向的 IS 元件结构单位，显然不论与中心区的相互方向如何，它们皆可发挥功能。对于复合型转座子来说，随着中央序列的增长，转座的频率降低。所以长度的依赖性是决定普通复合型转座子大小的一个因素。

支持复合型转座子转座的一个主要的动力是对中心区所携带的标志的选择。一个 IS*10* 结构单位自身可以随意地移动，并且比 Tn*10* 的移动频率高一个数量级。但对 tet^r 的选择可使 Tn*10* 维持在一起，因此在选择的条件下，完整的 Tn*10* 的转座频率可明显增加。

IS 元件所编码的转座酶的活性负责识别一个靶部位和识别转座子的末端。只有转座子的末端可作为转座的底物。

第三节　转座发生的机制[9~27]

转座子插入到一个新的部位的通常步骤是：在靶 DNA 上制造一个交错的切口，然后转座子与突出的单链末端相连接，并填充缺口。交错末端的产生和填充解释了在插入部位产生靶 DNA 正向重复的原因。在两个链的切口之间的交错决定了正向重复的长度；

这样每个转座子靶部位的重复特征反映了参与切割靶 DNA 的酶的几何学特征。

一、三种不同类型的转座

所有形式的转座皆利用交错末端，但是，我们能区别出三种不同类型的转座子移动机制(图 10－5)：

(1) 复制型转座(replicative transposition)　在转座反应过程中，转座的元件被复制，因此转座的序列实际上是原来元件的一个拷贝。如图 10－5(1)所示，作为自身移动的一个部分，转座子被复制，一个拷贝仍保留在原来的位置上，而另一个则插入到一个新的部位，这样转座过程伴随着转座子拷贝数的增加。复制型转座涉及到两种类型的酶：一种是转座酶，它作用在原来的转座子的末端，另一种是解离酶(resolvase)，它作用于复制拷贝元件。与 TnA 有关的一组转座子只通过复制型转座移动。

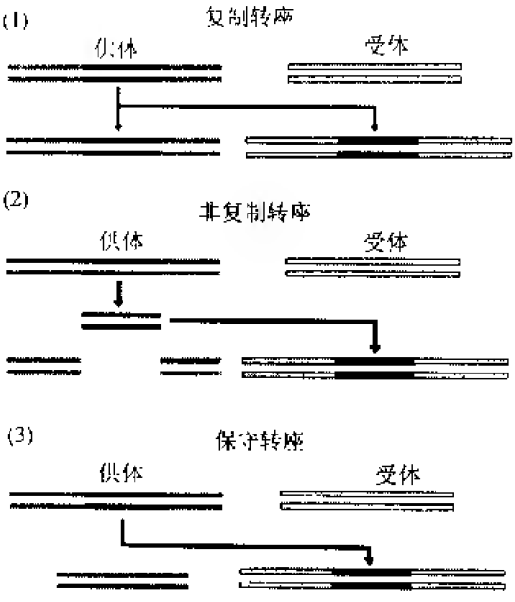


图 10－5　转座的三种形式
(1)复制转座　在复制转座时，先复制一个转座子拷贝，该拷贝插入到受体位点。供体位点不变，这样，供体和受体各有一个转座子的拷贝。(2) 非复制转座　转座子作为一个物理实体由供体位点移到受体位点，在供体位点留下一个断裂，如果不被修复可引起细胞死亡。(3)保守性转座　直接由供体部位移到受体部位。

(2) 非复制型转座(nonreplicative transposition)　转座元件作为一个物理实体直接由一个部位转移到另一个部位。这可以通过两种机制进行，一种利用供体和靶 DNA 序列的连接，并且有些步骤与复制型转座相同。插入序列和复合型转座子 Tn*10* 和 Tn*5* 则利用图 10－5(2)所示的机制，在该过程中转座子由与其相连的供体 DNA 释放，这种类型的机制只需要转座酶。非复制性转座的两种机制皆使转座元件离开供体部位并插入到靶部位。在非复制型转座发生之后，供体分子怎么样了呢？一种可能是供体即被破坏(这大概可以被具有多于一个染色体拷贝的细菌所耐受)，另一种可能是宿主修复系统识别了断裂的双链并修复了它[图 10－5(2)]。

(3) 保守型转座(consertive transpositon)　保守型转座描述另一种非复制型转座过程，在该过程中转座元件从供体部位被切除并通过一系列过程插入到靶部位，在该过程中每个核苷键皆被保留[图 10－5(3)]。这完全类似于 λ 的整合机制，并且此类转座子的转座酶与 λ 整合酶家族有关。应用这种机制的转座子是大的，并且除了能介导转移其自身外，还能使供体 DNA 由一种细菌转移到另一种。虽然这种元件最初被列为转座子，但它们更适合被看作是附加体。

虽然一些转座子只利用一种类型的转座途径，但也有些转座子可能利用两种途径。IS*1* 和 IS*903* 元件可利用非复制型和复制型两种转座途径，而 Mu 噬菌体可利用任一种转座途径，但先形成一个相同的中间物。

除了引起在新部位序列插入的简单分子间转座外，转座子还能促进其他类型的 DNA

重排，其中的某一些是多拷贝转座子间相互作用的结果，而其他的则代表了转座过程的另一些结果，并且它们留下了相关这些过程性质的一些线索。

当一个转座子在靠近它原来部位的第二个位点插入一个拷贝时，即有可能引起宿主DNA发生重排。宿主系统可能会使这两个拷贝之间发生交互重组，其结果由重复序列是相同方向或相反方向而定。

如图10-6(1)所示，任何一对正向重复序列之间重组会导致它们之间的序列发生缺失，这一中介区作为一个环形DNA被切除(从细胞内丢失)，染色体保留正向重复序列的一个拷贝。注意复合型转座子Tn*9*的正向重复序列IS*1*之间的重组可使该转座子被一个单一的IS*1*所替代。

因此，与一个转座子相连接部位的序列的缺失由两个阶段组成：转座产生一个转座子的正向重复序列，而正向重复序列之间发生重组。然而，出现在转座子附近的大部分缺失可能是由于转座途径之中和转座过程之后的变化所引起的。

图10-6(2)所示为一对反向重复序列之间交互重组的结果。重复序列之间的区域被倒位；重复序列本身被保留[图10-6(2)]，并有可能引发进一步的倒位。注意：虽然通过重组，复合型转座子中心区的方向被倒位，但它本身是基因组内的一个稳定组成部分。

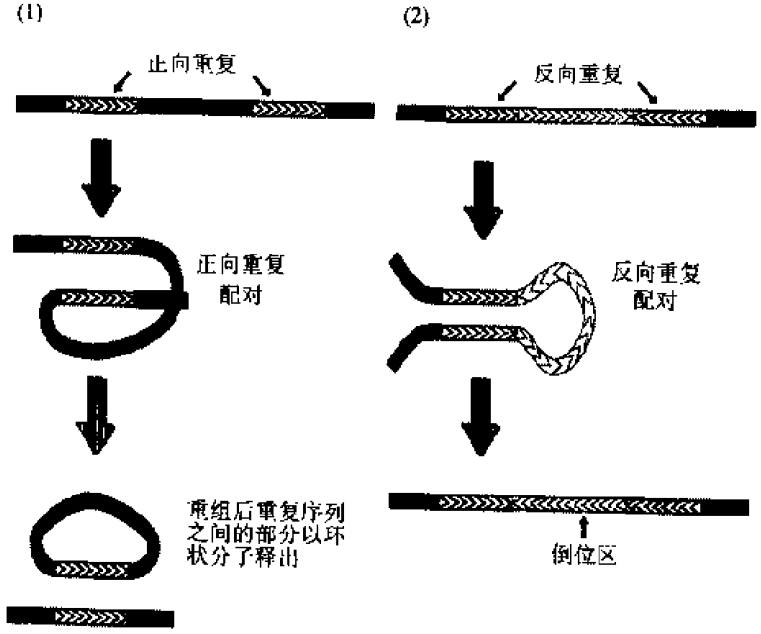


图10-6 交互重组模型

(1)在正向重复序列之间的交互重组，切除正向重复序列之间的部分，每一个重组体含一个正向重复序列的拷贝。

(2)在反向重复序列之间的交互重组，结果为反向重复序列之间的部分被倒位。

某些转座子促进复制倒位(duplicative inversions)。它们可通过一些转座子被鉴定，这些转座子的中心区与它通常的方向相反，而中心区的两侧为反向重复。

转座子本身不支持切除反应，但是当转座子的同源区被细菌的酶所识别时，切除即有可能发生。这一过程很重要，因为转座子的丢失有可能使原位点的功能得以恢复。精确的切除(precise excison)需要除去一个转座子，加上复制序列的一个拷贝，这是罕见的，对Tn*5*来说其发生的频率约为10^{-6}，Tn*10*为10^{-9}，它可能涉及到9 bp复制的靶部位之间

的重组。

不精确的切除(imprecise excision)留下一个转座子的残留物,虽然残留物可以有效地阻止靶基因的再活化,但可能不足以引起相邻基因的极性效应,因此会发生表型的改变。对 Tn*10* 来说,不精确切除发生的频率约为 10^{-6}。它涉及到 IS*10* 中的 24 bp 序列之间的重组,这些序列为反向重复序列,但因为 IS*10* 本身是反向的,因此在 Tn*10* 中它们形成正向重复。

同精确切除相比,不精确切除的发生频率较大,这可能与正向重复序列长度的增加有关(24 bp 对 9 bp)。两种切除类型都不依赖于转座子编码的功能,但其机制还不清楚。切除不依赖 RecA,是通过某种细胞机制在很靠近的重复序列之间产生自发缺失。

二、转座的共同中间体

许多可移动的 DNA 元件通过一种基本类似的机制由染色体的一个部位转座到另一个部位。它们包括 IS 元件、真核和原核的转座子和 Mu 噬菌体。反转录病毒 RNA 的 DNA 拷贝的插入也利用一个类似的机制,免疫球蛋白重组的第一阶段与其类似。

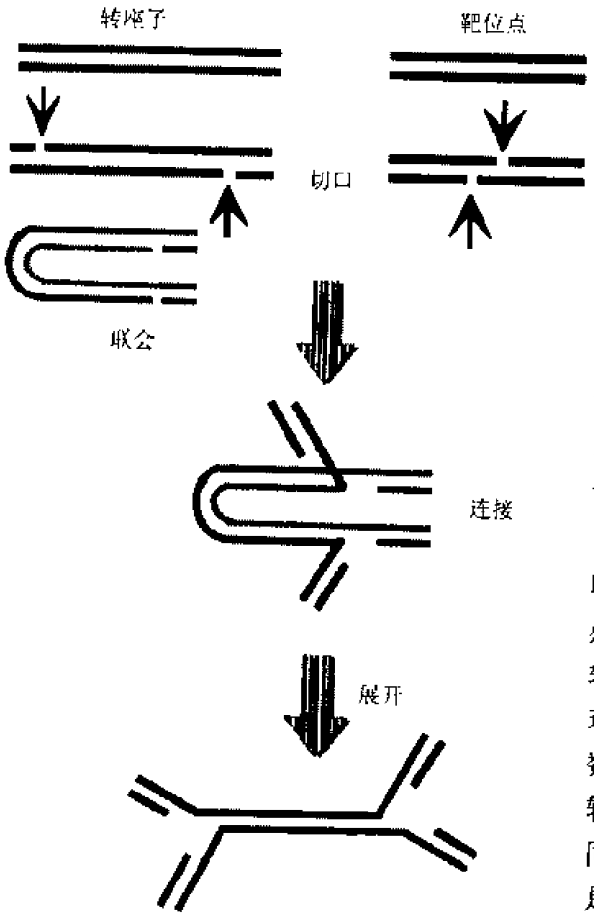


图 10-7 转座发生的机制
转座过程由转座子末端和靶位点的切口引发,然后连接切口末端形成链转移复合物。

转座起始于一个共同的机制,即是转座子与其靶部位相连接的机制。如图 10-7 所示,转座子在两端被切开,而靶部位的两个链也被切开,切口的末端被交叉相连,这使转座子和靶部位之间产生共价连接。在该过程中转座子的两个末端被带到一起,图中为了简单起见,把联会期放在切割之后,但实际上它发生在切割之前。

我们对这一过程了解最多的是 Mu 噬菌体的转座,它利用两种方式进行,当感染一个宿主细胞时,Mu 通过非复制型转座整合到宿主基因组中,在确保裂解循环(ensuring lytic cycle)的过程中,其拷贝数通过复制型转座而扩增。两种类型的转座皆涉及到转座子和其靶部位之间的同样类型的反应,但是,接着进行的反应是不同的。

噬菌体 DNA 转座的起始步骤通过 MuA 转座酶进行。Mu DNA 的每一末端皆各含一个 22 bp 共有序列的 3 个 MuA 结合位点,L1、L2、L3 位于左末端,R1、

R2、R3 位于右末端，MuA 的单体能与每个位点相结合，MuA 也能与噬菌体基因组的一个内部位点相结合。MuA 与左右两侧末端和内部位点相结合形成一个复合物。内部位点的作用还不清楚，它似乎对形成复合物是必需的，但与链的切割和随后的步骤无关(图 10-8)。

这只涉及到最靠近转座子每个末端的两个位点。MuA 亚单位与这些部位结合形成一个四聚体，这使转座子的两个末端形成联会。此四聚体可确保 Mu DNA 的两个末端以一种协调的方式反应。MuA 有两个对 DNA 进行操作的活性位点，而它们作用的模式使转座酶亚单位以反式进行作用。酶的共有序列结合位点与 L1、L2、R1、R2 的 22 bp 序列相结合。酶的活性位点在与 MuA 结合位点 L1 和 R1 邻近的部位切割 Mu DNA 链。但酶的活性位点不能切与该酶相结合侧的 DNA 共有序列。然而，它能以反式作用在不同 DNA 链上的适当部位切割。

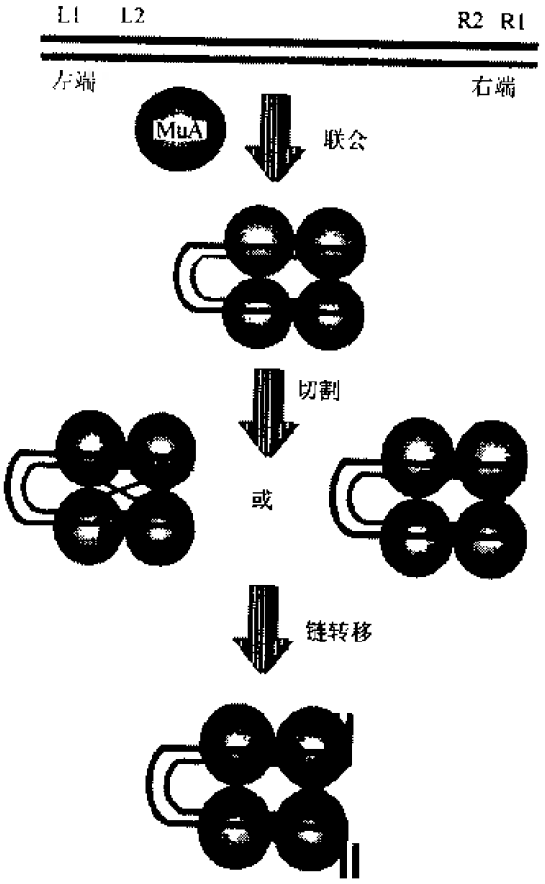


图 10-8 Mu 转座子 DNA 与靶部位连接的三阶段
MuA 转座酶在 Mu 噬菌体的末端联会处形成一个四聚体；转座酶的亚单位以反式作用切割 DNA 的末端；然后通过第二个反式作用与靶 DNA 的切口末端相连接。

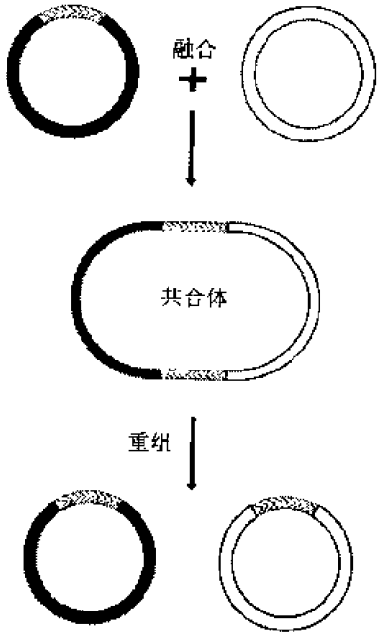


图 10-9 复制转座形成共合体
转座可融合供体和受体的复制子形成一个共合体，拆分后释放两个复制子，每个含有一个转座子拷贝。

这样，转座子的末端被 MuA 亚单位以反式作用的方式切割。反式的作用模型意味着实际上与 L1 和 R1 结合的单体不切割相连的部位，与左末端相结合的一个单体切割右末端的位点，反之亦然。(我们不知道在反应的这个阶段哪个单体是有活性的)。链转移

反应也以反式进行，位于 L1 的单体转移 R1 的链，反之亦然。有可能不同的单体催化一个已知末端的切割和链转移反应。

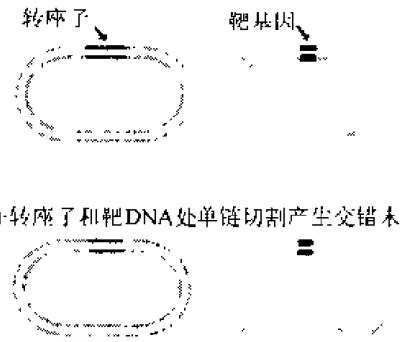


图 10－10　Mu 转座产生一个交叉结构，该结构通过复制转变为一个共合体

这些反应的产物是一个链转移复合物，在此复合物中转座子通过每个末端的一条链和靶位点相连。随后反应的差异性决定转座类型。链转移复合物可为复制（导致复制转座）或修复（导致非复制转座）的靶部位。

三、复制型转座的基本过程

1）复制型转座的基本途径见图 10－10。链转移复合物的 3′端被用作复制的引物，这就产生一个被称为共合体（cointegrate）的结构，它代表原来两个分子的融合体。共合体在原来两个分子之间结合处含有两个转座子的拷贝，方向为正向重复，如前部分所描述，交叉通过转座酶的作用而形成，它转变为共合体需要宿主的复制功能。

2）转座子两个拷贝之间的同源重组释放出两个复制子，每个复制子皆有一个转座子的拷贝，复制子之一是原来供体的复制子，另一个是靶部位的复制子，它获得了一个转座子，并且此转座子与宿主靶序列短的正向重复序列相连接。重组反应被称为解离反应，与其相关的酶被称为解离酶。

涉及到产生一个共合体的反应已在 Mu 噬菌体的研究中被详细地描述（图 10－10）。该过程以链转移复合物的形成开始（有时也称为交叉复合物），供体和受体的链相连接，即为转座子序列的两个末端与靶部位产生的两个单链突出末端相连接。链转移复合物产生一个交叉形的结构，此结构在双螺旋转座子处被连在一起，交叉结构的命运决定转座方式。

复制转座的原理是转座子通过转座被复制，使靶部位和供体部位皆含转座子的拷贝，其产物是一个共合体。

交叉结构在每一个交错的末端含有一个单链区，这些区为假复制叉，它提供 DNA 合成的模板（用作复制的引物末端这一事实表明链断裂必须是极性的，也就是在相应点上产生 3′－OH 末端）。

如果复制由两个假复制叉进行，它将通过转座子分开其双链，然后在末端结束。复制可能是通过宿主的复制系统来完成。此时该结构已变为一个共合体，在每个复制子之间的连接处为正向重复的转座子（如果沿着共合体序列追踪转座子，即可发现它们为正向重复序列）。

四、非复制型转座

交叉型结构也可被用于非复制型转座。通过这种机制的非复制型转座的原理为断裂和再结合反应使靶部位得以重构，供体仍然是断裂的，没有共合体形成。

Mu 噬菌体产生非复制型转座的切割过程，见图 10－11。一旦没有断裂的供体链产生切口，转座子两侧与靶部位的链即能被连在一起。通过交错切割产生的单链区必须通过修复合成反应所填充。反应的结果为转座子被插入到靶复制子中，并且其两侧为由原来的单链切口所产生的重复序列。供体复制子在原来转座子存在的部位产生一个双链断裂。

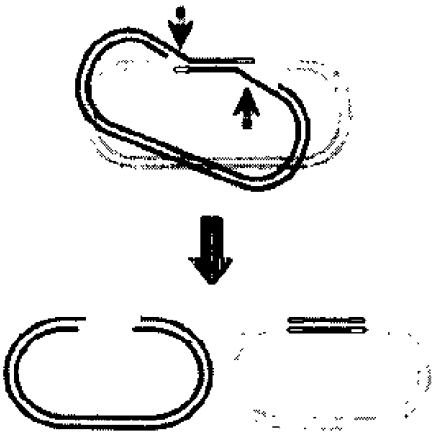


图 10－11　非复制转座模式
当交叉结构通过切割被拆分时产生非复制性转座。转座子插入到靶 DNA，两侧为靶 DNA 的正向重复序列，供体部位留下一个缺口。

非复制型转座也可通过另一种途径产生，在该过程中，在靶部位处产生单链切口，但在转座子两端产生双链断裂。转座子从供体的序列中被释放出来[图 10 5(1)]。Tn*10* 利用这种切割和连接过程进行转座(图 10－12)。

证明 Tn*10* 转座是非复制型的一个简捷的实验是利用一个人工构建的异源双链 Tn*10*，在该转座子中含有一个错配的碱基对。如果转座涉及到复制，在新部位的转座子将只含有来自一个亲代 Tn*10* 链的信息。但如果转座是通过转座子的物理移动产生，错配的碱基对将保留在新的部位上，转座后，通过检查转座子两条链所保留的信息即能证明错配是否保留。

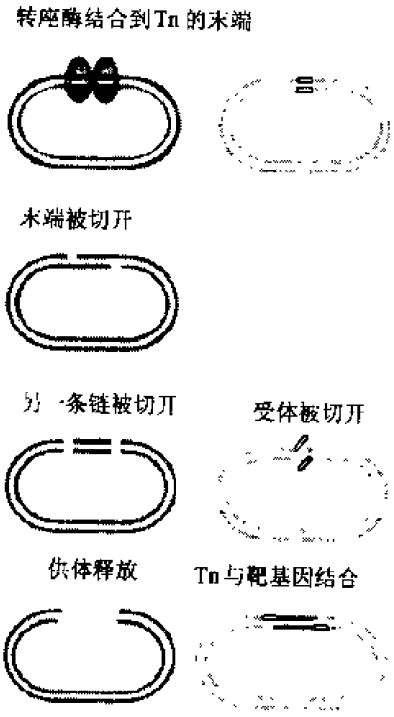


图 10－12　Tn*10* 非复制转座模式
依次切开 Tn*10* 的两条链，然后转座子与切开的靶部位相连接。

图 10－11 和图 10－12 模型的基本差别是在图 10－12 模型中，Tn*10* 在与靶部位进行任何连接之前其两条链即被切断。反应的第一步是转座酶识别转座子的末端，并形成一个蛋白质的结构复合物，反应发生在此结构内部。在转座子的每个末端，双链依一个特定的顺序被切割，首先被转移的链(与靶部位相连接的链)被切割，然后另一条链(与图 10－10、图 10－11 中 Mu 转座过程中的顺序相同)被切割。被切割的供体 DNA 此时被释放，然后转座子与靶部位的切口末端相连接。转座子和靶部位仍处在通过转座酶和其他蛋白质所产生的蛋白质结构复合物的控制之下，在转座子末端的双链切割阻止了任何复制型转座并迫使反应通过非复制型转座进行。其反应结果与图 10－11 所示相同，但切割和连接步骤的顺序不同。Tn*10* 转座酶以二聚体

的形式发挥功能，每个单体有一个活性部位，并相继催化转座子两条链的断裂和靶部位交错切口的形成

五、TnA 的转座

TnA 家族由大约 5 kb 长的大的转座子组成，复制型转座是其移动的唯一模型。它们不是依赖 IS 的复合式转座子，而是由独立的转座基因和特征性基因（如药物抗性基因）所组成。

TnA 家族包括几个相关的转座子，其中 Tn*3* 和 Tn*1000*（以前称 γδ）已被充分定性。它们有紧密相关的末端反向重复序列这一通常特征，末端重复序列的长度一般大约为 38 bp。任意一个重复序列内的顺式作用缺失将阻止转座子的转座。转座过程中，在靶部位产生一个 5 bp 的正向重复序列，它们携带着例如 amp^r 等抗性标志。TnA 介导的转座的两个阶段通过转座酶和解离酶完成。*tnpA* 和 *tnpR* 两个基因已通过缺陷型突变得到鉴定。像 IS 元件一样，转座过程涉及到转座子的末端。解离需要一个特异的内部位点，这是 TnA 家族的特有特征。

tnpA 的突变体不能转座。该基因的产物是一种转座酶，该酶与 38 bp 反向末端序列内的一个 25 bp 序列相结合。与转座酶结合位点相连的是大肠杆菌 IHF 蛋白的结合位点；转座酶和 IHF 以协同的方式与转座子末端相结合。转座酶识别靶 DNA。并且在该部位产生一个 5 bp 的交错断端。IHF 是一种 DNA 结合蛋白，在大肠杆菌中，该蛋白常常和组装大的结构有关；它在转座反应中的作用可能不是主要的。

tnpR 基因产物有双重功能，这是通过突变的多种效应所发现的。该蛋白可作为一种基因表达的抑制子并具有解离酶的功能。

tnpR 的突变可增加转座频率。这是因为 *tnpR* 同时抑制了 *tnpA* 和它自身基因的转录。因此，TnpR 蛋白的失活使 TnpA 合成增加，导致转座频率提高。这表明 tnpA 转座酶的量一定是转座的一个限制性因素。

tnpA 和 *tnpR* 基因从一个富含 A-T 的顺反子间控制区起始，以两个相反方向被转录（如图 10－13 中 Tn*3* 所示）。tnpR 的两个效应皆通过它在该区内的结合而介导。

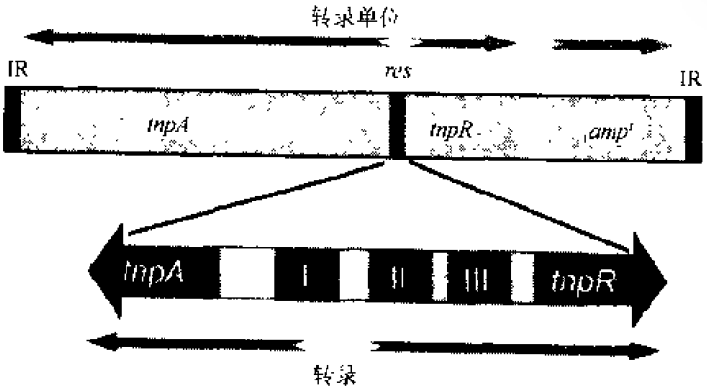


图 10－13　TnA 家族转座子的结构
TnA 家族的转座子有反向末端重复序列、一个内部的 *res* 位点和 3 个已知的基因。

作为解离酶，TnpR 与共合体结构中的 Tn*3* 正向重复序列之间的重组有关。原则上，在一个共合体中，转座子的两个拷贝间，任何相应的位点间的同源性重组皆能解离此共合体，但 Tn*3* 解离反应只发生在一个特异部位。

解离位点称为 *res*（resolvation site）。它是通过阻止转座

完成，并引起共合体堆积的顺式作用缺失而被鉴定的。在缺乏 *res* 时，解离反应能够被 RecA 介导的一般重组反应替代，但其效率很低。

TnpR 解离酶的结合部位见图 10－13 下部，共有三个结合位点，每个长约 30～40 bp，结合在每个位点上独立地发生，三个结合位点具有序列同源性，带有一个双重对称的共有序列。

位点 I 包括的区域一般被定义为 *res* 部位；当它缺失时，解离反应根本不能进行。然而，解离反应也与位点 II 和 III 的结合有关，这是因为如果这两个部位中的任何一个发生缺失，解离反应则只能很低效地进行。部位 I 包括 *tnpA* 转录的起始点。部位 II 包括 *TnpR* 转录的起始点；一个操纵基因突变部位恰恰位于该部位的左端。

这些部位之间相互作用吗？一种可能是所有三个部位的结合，可使 DNA 维持一个适当的拓扑学构象。与单一部位的结合可抑制 *tnpA* 和 *tnpR* 的转录，但不引起 DNA 的任何变化。

一个类似共合体的 DNA 分子可作为底物在体外进行解离测定，底物必须是超螺旋的；它的解离产生两个链环，每一个含有一个 *res* 位点，该反应需要大量的 TnpR 解离酶，但不需要宿主的因子。解离发生在一个大的核蛋白结构中，解离酶与每一个 *res* 位点相结合，然后结合部位被带到一起形成一个直径约 10 nm 的结构。在反应过程中超螺旋发生变化，由于转座酶的结合，DNA 在 *res* 位点发生弯曲。

解离是一个非复制反应，键的断裂和再连接不需要能量的输入，在共合体的解离中，一个中间复合物可被识别，它是由解离酶与在 *res* 位点进行的双链切割形成的两个 5′端共价结合而组成的。切割在一个短的回文区内对称地发生，并产生两个碱基的黏性末端。对位于位点 I 区的进一步的研究使我们可以描述切割反应如下所示：

5′TTATAA3′
3′AATATT5′

↓

5′TTA-T　　蛋白-AA3′
3′AA-蛋白　+　TATT5′

反应类似于 λInt 在 *att* 位点的作用，*res* 位点 20 bp 中的 15 个与 *att* 相应位置上的碱基相似。我们可依据切点排序如下：

res　GATAAT↓TTATAATAT
att　GCTTTTTTAT↑ACTAA

两个反应对 DNA 的操作过程是类似的，虽然解离只见于分子内位点之间，而 *att* 位点之间的重组是分子间的并且具有方向性。但是，在每种情况下蛋白质作用的机制不同，解离酶作用方式是以 4 个亚单位的形式结合于重组的 *res* 位点，每个亚单位进行一次单链的切割。随着亚单位之间相互的重新组合物理地移动 DNA 链，置它们于重组构象，然后切口被封闭。

六、Tn*10* 的转座

转座频率的控制对细胞来说是重要的。为了自身的生存，一个转座子必须维持某种

最低的移动频率，太高的频率可能损伤宿主细胞。每个转座子似乎皆有一个控制自身转座频率的机制。对 Tn*10* 来说，多种控制机制已得到鉴定。

Tn*10* 是一个复合型转座子，它的 IS*10R* 元件提供转座活性的组件。IS*10R* 的结构见图 10－14。靠近外侧边缘有两个启动子，启动子 P_{IN} 负责 IS*10R* 的转录，启动子 P_{OUT} 使转录向相邻的侧翼宿主 DNA 的方向进行，转录常常在转座子内终止，但偶尔进行到宿主的 DNA；有时这种通读式转录与活化相邻的细菌基因有关。

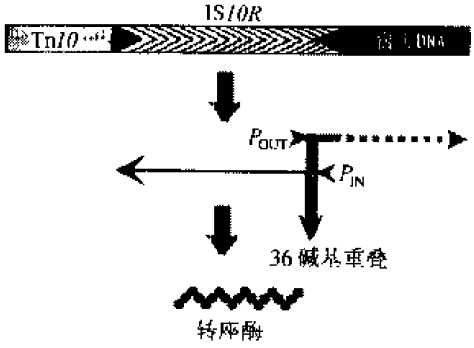


图 10－14　Tn*10* 复合型转座子 IS*10R* 元件的结构和转录

位于 IS*10R* 的边界处附近有两个反向排列的启动子，强启动子 P_{OUT} 向相连的宿主 DNA 的方向转录。弱启动子 P_{IN} 产生一个 RNA 转录物，该 RNA 转录物转录至 IS*10R* 的长度并经翻译产生转座酶。

“多拷贝抑制”(multicopy inhibition)的现象揭示了 IS*10R* 转座酶基因的表达是受调控的。当通过一个质粒引入多个 IS*10R* 拷贝时，细菌染色体上 Tn*10* 转座子的转座即相应减少。抑制需要 P_{OUT} 启动子，并且在翻译水平上进行。其作用的基础是 P_{IN} 和 P_{OUT} 转录物 5′末端区含有碱基的重叠。*OUT* RNA 是一个 69 bp 长的转录物，它的存在水平大于 *IN* RNA 100 倍，其原因有两个：P_{OUT} 是一个比 P_{IN} 强的启动子，并且 *OUT* RNA 比 *IN* RNA 更稳定。

OUT RNA 作为反义 RNA 发挥其功能。在单拷贝的情况下，*OUT* RNA 没有影响，但当拷贝数大于 5 时，即有明显的影响。通常每一个 IS*10* 拷贝对应于大约 5 个 *OUT* RNA 拷贝(在一个典型的多拷贝的情况下，其对应的 *OUT* RNA 为 150)。*OUT* RNA 碱基同 *IN* RNA 碱基配对；在与核糖体接触之前，过量的 *OUT* RNA 可确保 *IN* RNA 快速地与其结合，使配对的 *IN* RNA 不能被翻译。

转座酶蛋白的量常常是一个重要的特征。Tn*10* 转座酶只以每代每个细胞 0.15 个分子的低水平合成，Tn*10* 显示几个有趣的调控机制，图 10－15 所示为不同的作用对转座频率的影响。

在 IS*10R* 的一条链上，有一个编码转座酶的连续可读框。转座酶的水平限制转座的频率。该基因的突变子可通过另一个野生型 IS*10* 元件以反式作用方式补充，但有一些

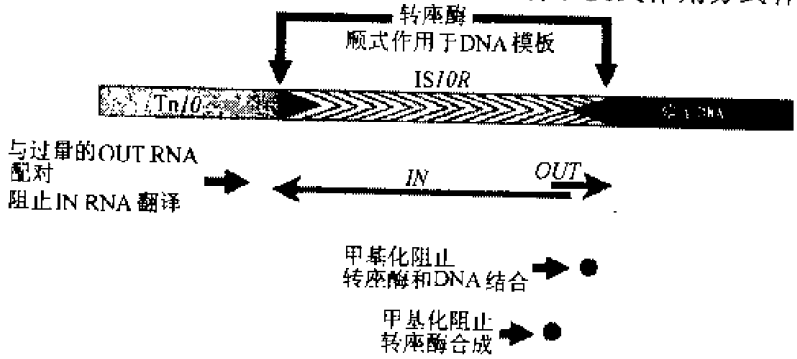


图 10－15　影响转座频率的机制

通过影响转座酶蛋白合成的几种机制抑制 Tn*10* 的转座频率。发生在复制后的甲基化可限制转座子的转座。在多拷贝的情况下，顺式作用倾向限制靶部位的选择，而 *OUT*/*IN* RNA 配对抑制转座酶的合成。

困难。这反映出转座酶对顺式作用有较强的倾向性，该酶只能有效地作用于通过转录和翻译产生的它自身的DNA模板。顺式作用的倾向是由IS元件编码的转座酶的共同特征（另一些倾向顺式作用的蛋白包括和 $\Phi X174$ 复制有关的A蛋白）。转座酶的顺式作用倾向可限制转座的频率，它可确保在细胞中有较多的Tn*10*拷贝存在的情况下转座酶的有效浓度不随之增加。

顺式作用倾向反映了转座酶能更有效地识别那些与酶的合成部位更近的DNA靶序列的能力吗？一种可能的解释是在蛋白合成后（甚至在合成时）转座酶与DNA非常紧密地结合以致于它很少有可能扩散到别处去。另一种可能性是当酶不与DNA相结合时，它可能是不稳定的，这样不能很快与DNA结合的蛋白质永远不会有变为活性蛋白质的机会。

顺式作用倾向和多拷贝的抑制共同作用的结果可确保在一个细菌基因组内，Tn*10*拷贝数目的增加不会引起可导致损伤基因组的转座频率的增加。

dam 甲基化的作用为每个元件提供了最重要的调节系统。它减少转座频率，并且（更重要地）使复制和转座偶联。IS*10* 转座的能力通过转座子对两个位点的甲基化状态的反应与复制循环相关联。一个位点位于IS*10R* 末端的反向重复内，转座酶结合在此处，另一个位点是位于 P_{IN} 启动子内，转座酶基因由此被转录。

这两个位点皆通过 *dam* 系统被甲基化，Dam甲基化酶修饰通过复制产生的新DNA链上GATC序列的腺嘌呤。在 *dam*$^-$ 菌株内，两个靶部位缺乏甲基基团，Tn*10* 转座的频率提高1000倍。

在DNA复制时，新合成的GATC中的腺嘌呤不立即被甲基化，这样，在新复制的两个拷贝中，只有一条链上的位点被甲基化，产生半甲基化序列，在没有发生甲基化的链上由 P_{IN} 起始的转座酶基因的转录频率提高并且使转座酶与IS*10R* 末端的结合力增强，这两种作用相结合，共同活化了转座子。在野生型细菌中，这些位点在复制后一个短时期内仍保持半甲基化状态。

为什么需要在复制后不久即发生转座呢？Tn*10* 的非复制转座机制置供体DNA于被破坏的危险之中[图10-5(2)]。如果复制刚好产生供体序列的第二个拷贝，则细胞的存活机会可能会增加。因为两个新复制的拷贝只有一个可发生转座，所以这一机制是有效的（这是因为在转座的那条链 *dam* 位点是非甲基化的有关）。

因为一个转座子的靶部位的选择是随机的，所以有相当的可能性，它可转座到一个活性的操纵子内。是否有可能外部的转录能通过转座子并活化转座酶，并且转座酶的过量产生反过来又导致高水平的转座（可能是致死的）。Tn*10* 通过两种机制保护其自身免于这种过程，首先，通过IS*10R* 末端的转录降低它的活性，也有可能是通过抑制它结合转座酶的能力而实现的。另外，延伸于启动子上游序列的mRNA是很少被翻译的，这是因为它有一个二级结构使起始密码子难以接近。

第四节　玉米的转座元件[41~42]

一、玉米的控制元件

起始于40年代的玉米遗传学的研究表明，玉米的基因组在体细胞分裂时可发生变

化，这些变化是由控制元件所引起的，它们能由一个部位转移到另一个部位。控制元件最先由McClintock用遗传学方法进行了鉴定，现在被确定是一种转座元件，并可直接与细菌的转座元件相类比。

玉米的两个特征有助于转座过程的追踪，控制元件常常插入在表型上是可见的但又没有致死作用的基因附近。并且因为玉米展示克隆性发育，所以转座过程的发生和发生时间可以被观察到（图 10－16）。

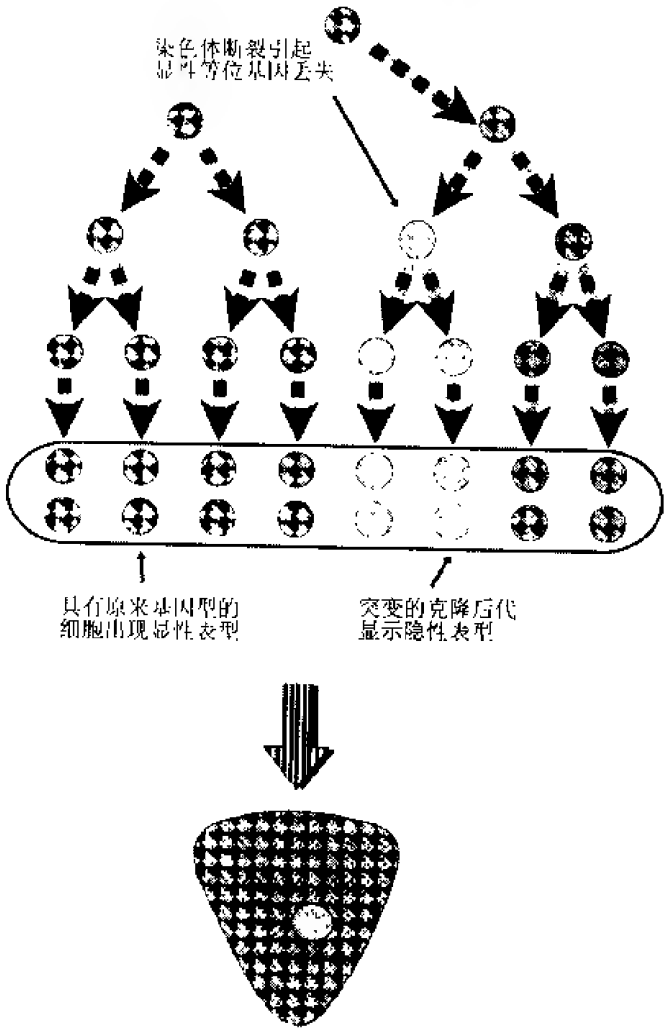


图 10－16　克隆分析鉴定了由一个单一祖先传下来的一组细胞。该祖先通过转座过程表型已发生改变。

该过程的性质并不重要，它可能是插入、切除或染色体的断裂，重要的是它发生在杂合子上并改变等位基因的表达。这样，经历这种过程的细胞的后代即会展现一种新表型，而未受影响的细胞的后代则继续显现原来的表型。

一个细胞有丝分裂的后代仍存在于同样的部位，这就可观察到一个表型变化的细胞组成部分。在体细胞发育过程中表型的改变被称为斑驳现象（variegation），其表现为在原有表型组织内的一部分细胞显示以新的表型。这一部分大小取决于它的谱系分裂的次数；这样新表型面积的大小即可通过基因型改变的时间长短而定。在细胞谱系中它出现的越早，其后代的数目就越多，在成熟组织中的体积也就越大。这可清楚地见于谷粒颜色的变化，即当一块新的颜色出现在另一种颜色内的时候。控制元件的插入可能会影响相邻基因的活性，在控制元件存在的那些部位，缺失、复制、倒位和易位都可能发生。染色体断裂是某些控制元件存在的常见结果。玉米系统的一个独特的特性是其发育过程中，控制元件的活性受到调控。在植物发育过程中，这些元件在特定的时间和频率发生转座并促进遗传重排。

在玉米中，控制元件特性的鉴定以 *Ds*（dissociation）元件为代表，该元件最初是通过它能提供染色体一个断裂位点这一特性而被鉴别（图 10－17）。考虑在一个杂合子中，*Ds* 位于着丝粒和一系列显性标志（*CI*、B_2 和 W_X）之间的同源染色体上，而另一个同源染色体上缺乏 *Ds*，并含有隐性标志（*c*、*bz* 和 *wx*）。在 *Ds* 处的断裂产生一个携带显性标志的

无着丝粒的片段，由于缺乏着丝粒，该片段在有丝分裂过程中丢失。这样后代细胞中只含有被完整染色体携带的隐性标志。

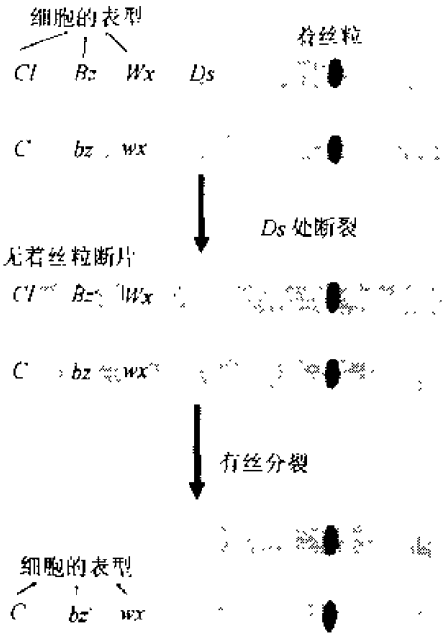


图 10-17 控制元件断裂的影响
一个控制元件的断裂引起无着丝粒的片段丧失；如果该片段携带一个杂合子的显性标志，它的丧失可改变表型。显性标志 $C1$、Bz、Wx 能通过细胞的颜色或通过适当的染色被观察到。

在 *Ds* 处的断裂可导致两个异常染色体的形成。它们是通过复制产物断端的连接而产生的。一个是U形的无着丝粒片段，它是由 *Ds* 远端的姐妹染色体连接而成。另一个是U形的双着丝粒染色体，由 *Ds* 近端的姐妹染色体连接组成。后者的结构导致典型的断裂-融合-桥循环。

当双着丝粒染色体在有丝分裂纺缍体上发生分离时，其结局如下，两个着丝粒被各自向相反的一极牵拉，这种力量使染色体在两个着丝粒之间的随机部位发生断裂。这种断裂-融合-桥循环随着细胞的进一步分裂而继续进行，这使子代继续发生遗传学的改变。

二、玉米的转座子家族

玉米基因组内含有几个控制元件家族，它们的数量、类型和位置在每个玉米株内皆不相同。每个家族的成员可被分成两种类型：

(1)自主元件(autonomous elements)　有切除和转座的能力。因为自主元件的持续的活性，在任何基因座，它的插入产生一个不稳定的或易变的等位基因。元件本身的丧失或者其转座能力的丧失，使一个不稳定的等位基因转化成为一个稳定的等位基因。

(2)非自主元件(nonautonomous elements)　它们是稳定的，不转座或在某种条件下发生自发性的改变。只有当同一家族的一个自主元件存在于该基因组的其他部位时，它们才变得不稳定。当通过一个自主元件反式互补时，一个非自主元件可展示同一个自主元件相关的活性范围，包括转座到新的位点的能力。非自主元件是自主元件由于丧失转座所需的反式作用的功能而产生的。

玉米控制元件家族是由一个单一类型的自主元件和许多不同的非自主元件所组成的。在一个家族中的非自主元件可通过自主元件以反式作用而激活。玉米中控制元件的主要家族见图 10-18。

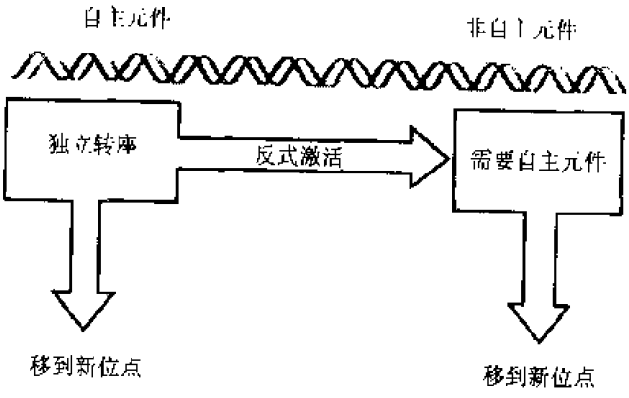


图 10-18 玉米控制元件的家族
玉米的控制元件家族分为自主和非自主元件两类，自主元件能进行转座。非自主元件缺乏转座的能力。自主元件和非自主元件能被分为4个以上的家族。

分于水平的鉴定表明，玉米的转

座子具有一个共同的结构特征，其末端为反向重复，并在靶 DNA 中产生短的正向重复序列，然而在其他方面如大小和编码能力上却不相同。经充分鉴定的家族是 *Ac*（activator element）元件和 *Spm* 元件。在一个植物的基因组中每个家族通常有几个成员存在（大约 10 个）。通过对 *Ac*/*Ds* 家族自主和非自主成员的分析，我们获得了这些元件许多个例的分子信息，它们的结构见图 10－19。

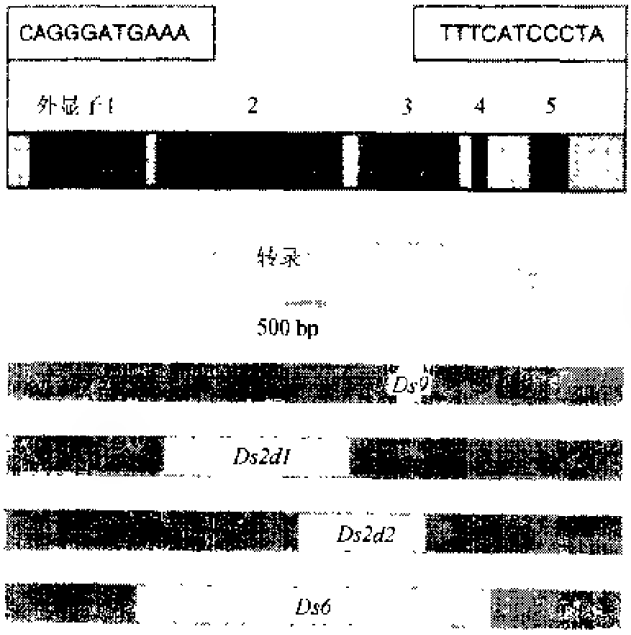


图 10－19　*Ac* 和 *Ds* 元件的结构
Ac 元件有两个可读框，*Ds* 元件有内部序列的缺失。

自主 *Ac* 元件序列的大部分被一个由 5 个外显子组成的单一基因所占据，其产物为转座酶。该元件本身末端为 11 bp 的反向重复序列；在插入位点，8 个碱基对的靶序列被复制。

Ds 元件各组成成员的长度和序列并不相同，但皆与 *Ac* 相关。它们有相同的 11 bp 反向重复末端。它们较 *Ac* 短，但缺失的长度有差异。在一个极端，*Ds*9 元件只有 194 bp 的缺失，而在另一极端，*Ds*6 元件只保留 2 kb 的长度，对应于 *Ac* 的每一端的 1 kb 区。在一个复合式双 *Ds* 元件中，一个 *Ds*6 序列反方向插入到另一个之中。

非自主元件缺乏内部序列，但具有末端反向重复序列（和可能的其他序列特征）。非自主元件是由自主元件通过缺失（或其他变化）而产生的，这些变化使反式作用转座酶失活，但却保留了完整的转座酶的作用位点（包括末端）。它们的结构包括含少量（无活性的）*Ac* 突变体到经过较大缺失或重排的序列。

另一个极端的例子，*Ds1* 家族成员由短的序列组成，它与 *Ac* 的相关性仅在于其末端存在反向重复序列。这类元件不需直接由 *Ac* 衍生而来，它们可能来自于产生反向重复序列的任何过程。它们的存在表明，转座酶只识别末端反向重复序列或可能某种短的内部序列与末端重复序列一起被识别。

Ac/*Ds* 的转座通过非复制型机制发生。克隆分析表明，*Ac*/*Ds* 的转座几乎总是发生在供体元件被复制之后。这些特征类似于细菌转座子 Tn*10* 的转座。受体部位常常与供体部位位于同一染色体上，并常常相当接近。

复制产生一个可能的 *As*/*Ac* 供体的两个拷贝，但通常只有一个拷贝转座。在转座后供体部位的命运如何？供体部位由于转座子的丧失而发生的重排可通过染色体的断裂来解释（图 10－17）。

自主和非自主元件在它们所处的环境下可发生各种变化。其中一些变化是遗传性的，另一些是非遗传性的。

当然，主要的变化是一个自主元件转变成非自主元件，但非自主元件也可发生进一步的变化。顺式作用的缺陷可使一个非自主元件不受自主元件的影响。因为它不再被活化

而转座，这种非自主元件有可能变得永久的稳定。

自主元件受“相的改变”的影响而发生变化，这些变化是可遗传的，但相对不稳定。

在植物的发育过程中，转座元件可以一种可逆性失活的方式循环于活性和无活性状态之间。因此相影响转座的时间和频率。

自主元件 *Ac* 和 *Mu* 相的变化似乎是由 DNA 的甲基化所引起的。活性和无活性元件对限制酶敏感性的比较表明，无活性元件在靶序列 $\frac{CAG}{GTC}$ 上被甲基化。每个元件中有几个靶位点，我们尚不知道哪些位点控制这种效应。我们同样想知道是什么控制这些元件的甲基化和去甲基化。转座可能是自我调控的，在基因组中，*Ac* 元件数量的增加使转座频率降低。*Ac* 元件可能编码转座的阻抑物；其活性可能是由提供转座酶功能的同一蛋白所提供。

三、*Spm* 元件影响基因表达

Spm 和 *En* 自主元件实际上是一样的，它们的差异不过 10 个位置（图 10－20）。重复序列对转座是很重要的，正如在末端缺失的转座缺陷型表型中所见到的那样，与 *Spm* 相关的转座子也见于其他植物，由于它们具有基本相似的结构而被划分为同一家族的成员。它们全部具有基本相同的末端反向重复序列，并在转座的靶 DNA 产生 3 个 bp 的正向重复序列。由于末端的相似性，它们被称为 CACTA 类转座子。

一个 8300 bp 的序列从该元件左侧末端的启动子处被转录，转录物中含有的 11 个外显子被剪接成一个 2500 bp 的 mRNA，该 mRNA 编码一个含 621 个氨基酸的蛋白质。该基因被称为 *TnpA* 基因，而该蛋白产物与该元件末端区的以多拷贝形式存在的 12 bp 共有序列相结合。TnpA 的功能是切割所需要的，但单独不足以完成切割。

这个家族所有的非自主元件（*dSpm* 代表缺陷型 *Spm*）与 *Spm* 元件本身在结构上是紧密相关的，它们含有影响 *tnpA* 外显子的缺失。

有两个可读框（ORF1 和 ORF2）位于 *tnpA* 第一个长的内含子内，它们包含在经另一种方式剪接的 6000 bp 的 RNA 中，它占 *tnpA* mRNA 的 1%。ORF1 和 2 称为 *TnpB*。它可能提供一种与 13 bp 反向重复末端相结合并为了转座而切割末端的蛋白质。

除了高活性的 *Spm* 元件外，还有 *Spm-u* 衍生物，它们在转座中显示较弱的活性。在图 10－20 的例子中，含有缺失 ORF1 和 ORF2 的活性 *Spm*，这表明在转座中对 *TnpB* 的需要可被绕过或替代。

Spm 的插入能控制插入部位基因的表达，受体基因座可被置于正或负的调控下，*Spm* 抑制型的基因座的表达受到抑制，*Spm* 依赖型的基因座只有在 *Spm* 的辅助下才被表达。当插入元件为 *dSpm* 时，抑制或依赖作用由自主型 *Spm* 以反式作用方式提供。这些相反作用的基础是什么呢？*dSpm* 抑制型的等位基因的外显子内含有插入的 *dSpm*，这就显产生一个问题，在外显子上含 *dSpm* 插入序列的基因如何能被表达呢？*dSpm* 序列能够通过它的末端序列从转录物中被切除。剪接过程可能在 mRNA 序列上留下一种变化，这就解释了它编码蛋白质特性上的变化。一些 *Ds* 插入序列亦具有类似的可由转录物中被切除的能力。

tnpA 提供 *Spm* 元件抑制基因的功能。缺陷元件的存在可以减低但不能完全阻止它

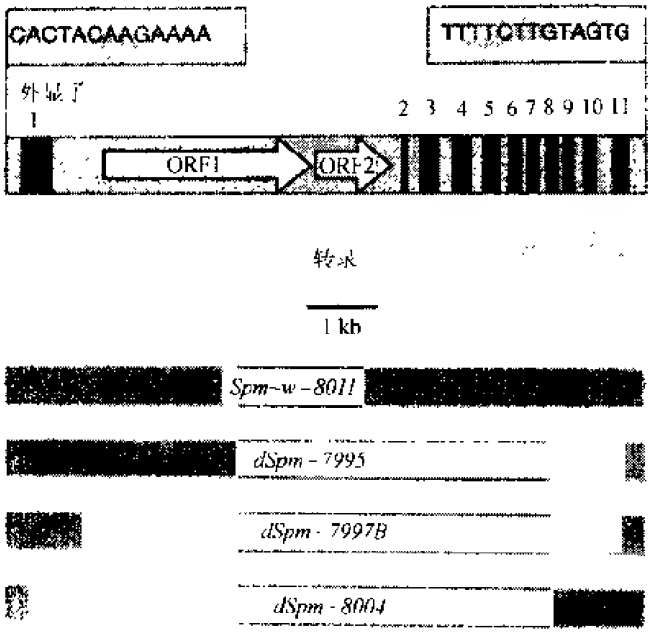


图 10-20 *Spm/En* 的结构

Spm/En 有两个基因，*tnpA* 由 11 个外显子组成，转录后经剪接形成含 2500 个碱基的 mRNA，*tnpB* 的 mRNA 含 6000 个碱基，有 ORF1 和 ORF2 两个可读框。

所插入的基因的表达。

然而，具有功能性 *tnpA* 基因的自主元件的引入可以完全抑制靶基因的表达。这是由于 TnpA 蛋白可与缺陷元件结合，阻断正在进行的转录，从而产生抑制作用。

一个 *dSpm* 依赖型的等位基因含有一个在基因附近而不是在基因内部插入的 *dSpm*。这种插入序列似乎可提供受体基因座--个活化该基因启动子的增强子。

dSpm 元件的抑制和依赖作用似乎皆依赖于自主型 *Spm* 元件 *tnpA* 基因反式作用的产物和该元件末端顺式作用的位点之间的相互作用。这样，取决于该元件位于受体基因的上游还是内部，蛋白和该元件末端之间的单一作用即可抑制或活化一个靶基因座。

Spm 元件的多种不同状态存在，其范围可由完全活性状态到隐性状态。隐性状态的元件是沉默的，它既不发生自身的转座也不活化 *dSpm* 元件。隐性元件可以被瞬间激活或通过与具完全活性的 *Spm* 元件的相互作用而转化成活性状态。灭活是由位于转录起始点附近序列的甲基化所引起的。目前还不清楚与甲基化灭活或去甲基化(或阻止甲基化)活化有关的反应的性质。

第五节 果蝇的转座元件[43~47]

研究表明当果蝇的某些品系进行种间杂交时即会遇到困难。当两个品系杂交时，后代即显示不育特性，其一系列的缺陷包括突变、染色体畸变、减数分裂中的异常分离和不育。这些相关的缺陷被称为杂种不育(hybrid dysgenesis)。

在果蝇中，与杂种不育有关的两个系统已被鉴定。第一个系统的果蝇被分成 I 型(inducer，I)和 R 型(reactive，R)，生育力降低的现象只见于雄性 I 和雌性 R 交配产生的后代，而不见于反向交配产生的后代。第二个系统中果蝇被分成 P(paternal contributing，P)型和 M(maternal contributing，M)型。雄性 P 和雌性 M 之间交配产生杂种不育，但反向交配时则不存在此种情况。

不育主要是与生殖细胞有关的一种现象。在 P-M 系统的杂交后代中，F_1 杂交果蝇具有正常的体细胞组织，但它们的性腺不发育。配子发育形态学上的缺陷可追溯到生殖细胞快速分裂阶段。

雄性 P 的任何一个染色体，在同雌性 M 交配时皆能诱导产生杂种不育。重组染色体

的构建显示，每个 P 染色体内皆有几个区域能引起不育。这表明，一个雄性 P 含有大量的 *P* 元件(P factors)，这些序列在染色体上占据许多不同的部位，这些部位随 P 株的不同而不同，*P* 元件不存在于 M 果蝇的染色体中。

在不育的个体中，诱导突变的有关过程是通过对 ω 突变体的 DNA 序列分析而鉴定的。所有的突变皆起因于 DNA 在 ω 基因座的插入。插入的序列被称为 *P* 元件，它们形成一个典型的转座系统。每个元件在长度上不同但在序列上是同源的，所有的 *P* 元件皆具有 31 bp 的反向末端重复序列，并在转座的靶 DNA 部位产生 8 bp 正向重复序列。最长的 *P* 元件大约 2.9 kb 长并有 4 个可读框。较短的 *P* 元件以相当经常的频率出现，很显然，它是完整的 *P* 元件的内部缺失而产生的，至少一部分短的 *P* 元件丧失了产生转座酶的能力，但它们可通过一个完整的 *P* 元件编码的转座酶以反式作用方式所活化。一个 P 品系含有 30～50 个 *P* 元件的拷贝，其中大约三分之一具有全长。M 品系中不存在这些元件。在 P 品系中，这些携带的元件是基因组的惰性成分。但是，当一个雄性 P 和雌性 M 交配时，它即变成具有转座活性的成分。

P-M 杂种不育的果蝇的染色体中，在许多新位点含有插入的 *P* 元件。杂种不育典型的染色体断裂发生的热点正是 *P* 元件存在的部位。*P* 元件转座到 M 染色体上的平均速率是大约每代一次。

P 元件的活化具有组织特异性：它只发生在性细胞之中。但 *P* 元件在性细胞和体细胞组织中均被转录。组织特异性是通过剪接类型的改变而获得的。

该元件的结构和转录见图 10－21。主要转录物为 2.5 kb 或3.0 kb 长，其差别可能仅反映对末端终止码的通读。可产生两种蛋白产物：

1) 体细胞仅前两个内含子被剪接，产生一个 ORF0-ORF1-ORF2 编码区。此 RNA 经翻译产生 66 kDa 的蛋白，该蛋白是转座活性阻抑物。

2) 在生殖细胞中，附加的剪接过程除去了内含子 3。这使所有 4 个可读框被连接成一个 mRNA，并经翻译产生一个 87 kDa 的蛋白，该蛋白为转座酶。

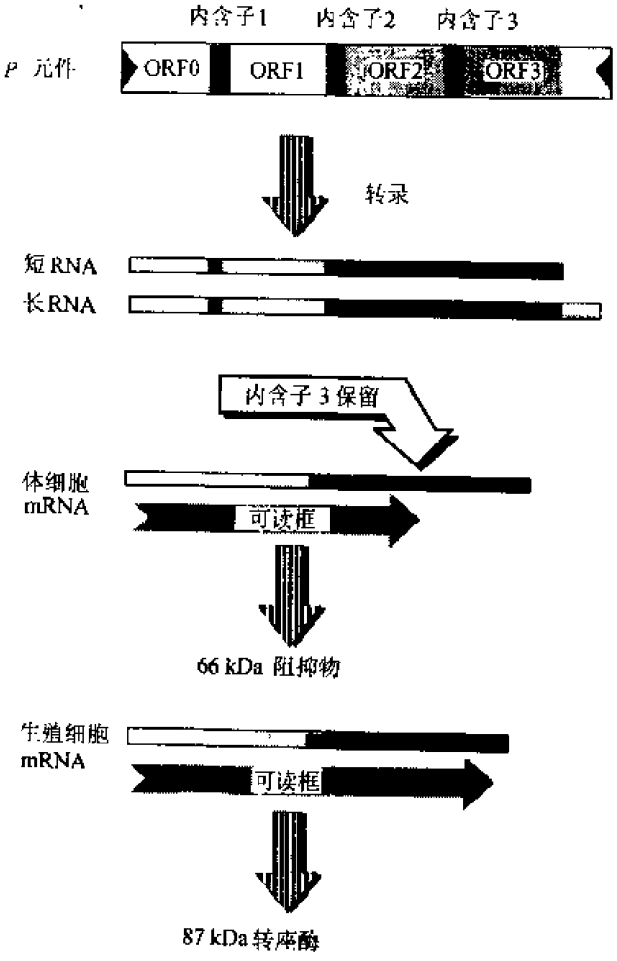


图 10－21　*P* 元件的结构和它的转录
P 元件有 4 个外显子，前 3 个被剪接后相连在体细胞内表达。全部 4 个外显子被剪接后相连在性细胞内表达。

两种类型的实验证实了第三个内含子的剪接是转座所需要的。第一，如果在体外使剪接连接部分发生突

变，然后将此 *P* 元件再引入果蝇，其转座活性消失。第二，如果第三个内含子缺失，这使 ORF3 被包括在所有组织的 mRNA 中，在这种情况下，转座既发生在生殖细胞之中也发生在体细胞之中。

因此，当 ORF3 经剪接而连入可读框时，*P* 元件即转成活性状态。这是关键的调控过程，并且它通常只发生在性细胞之中。

是什么决定组织特异性剪接呢？体细胞含有一种蛋白可与外显子 3 的序列结合并阻止最后一个内含子的剪接。在性细胞中缺乏这种蛋白，这使剪接后产生编码转座酶的 mRNA。

P 元件的转座大约需要 150 bp 的末端 DNA，转座酶与两个 10 bp 的序列相结合，此序列邻近于 31 bp 的反向重复序列。转座通过非复制型发生，其切割和连接机制类似于 Tn*10*。它以两种形式引起杂种不育，在新位点转座元件的插入可引起突变，而在供体部位遗留的断裂具有有害的作用[如图 10-5(2)模型所示]。

有趣的是在很大的比例上，断裂的供体 DNA 可通过同源染色体的序列而被修复。如果同源部位有一个 *P* 元件，则供体部位的 *P* 元件有可能被恢复(这类似于复制转座的结果)。如果同源部位缺乏 *P* 元件，修复产生缺乏 *P* 元件的序列，这就很明显地产生了精确切除(一个在其他转座系统中罕见的过程)。

杂种不育依赖于交配过程性的取向，这一事实表明胞浆与 P 因子本身是同样重要的。胞浆的作用被描述为细胞型，含 *P* 元件的果蝇品系为 P 细胞型，而缺乏 *P* 元件的果蝇品系为 M 细胞型。只有当含有 *P* 元件的染色体进入到 M 细胞型的胞浆中时才会产生杂种不育的现象，即当雄性亲代含有 *P* 元件，而雌性亲代没有时。

细胞型显示出一种胞浆的可遗传效应，当杂交通过 P 细胞型发生时(雌性亲代含有 *P* 元件)，其产生的杂交后代可连续几代与雌性 M 型杂交而不产生杂种不育。因此在 P 细胞型胞浆中有某种物质可抑制杂种不育经过几代后被稀释掉(细胞型的这种作用的分子模型见图 10-22)。它依赖于 66 kDa 蛋白抑制转座的能力。该蛋白作为卵子中的母体因子而存在。尽管存在 *P* 元件，在 P 株中一定有足够的蛋白阻止转座发生。在任何涉及雌性 P 的杂交中，该蛋白的存在阻止转座酶的合成或活化。但是当雌性亲代是 M 型时，在卵子中没有阻抑物，由雄性亲代引入的 *P* 元件在生殖细胞中产生转座酶的活性。P 细胞型可通过多代发生作用的能力表明在卵子中一定有足够的阻抑物存在，它一定是足够稳定的，并通过成年动物传到下一代的卵子中。

30 多年前采集的野生型果蝇传下来的果蝇品系总是 M 型，而从 10 年前采集的果蝇传下来的品系总是 P 型。这是否意味着 *P* 元件家族最近一些年来侵入到果蝇的野生型群体中呢？当 *P* 元件被引入到一个新的群体中时，确实是高度侵入性的。侵入的 *P* 元件应来源于另外一个品种。

由于杂种不育使品种间杂交减少，它是物种分化途径中的一个阶段。假设通过一种可转座的元件在某一部位产生一个不育的系统，而另一种转座元件可在别的部位产生一个不同的系统。果蝇在这两个区域由于这两个系统的存在将产生不育。如果这导致它们不育，群体在遗传上便会分离开来，可能会发生进一步分离。因此，多个不育系统的存在可导致不能交配和物种分化。

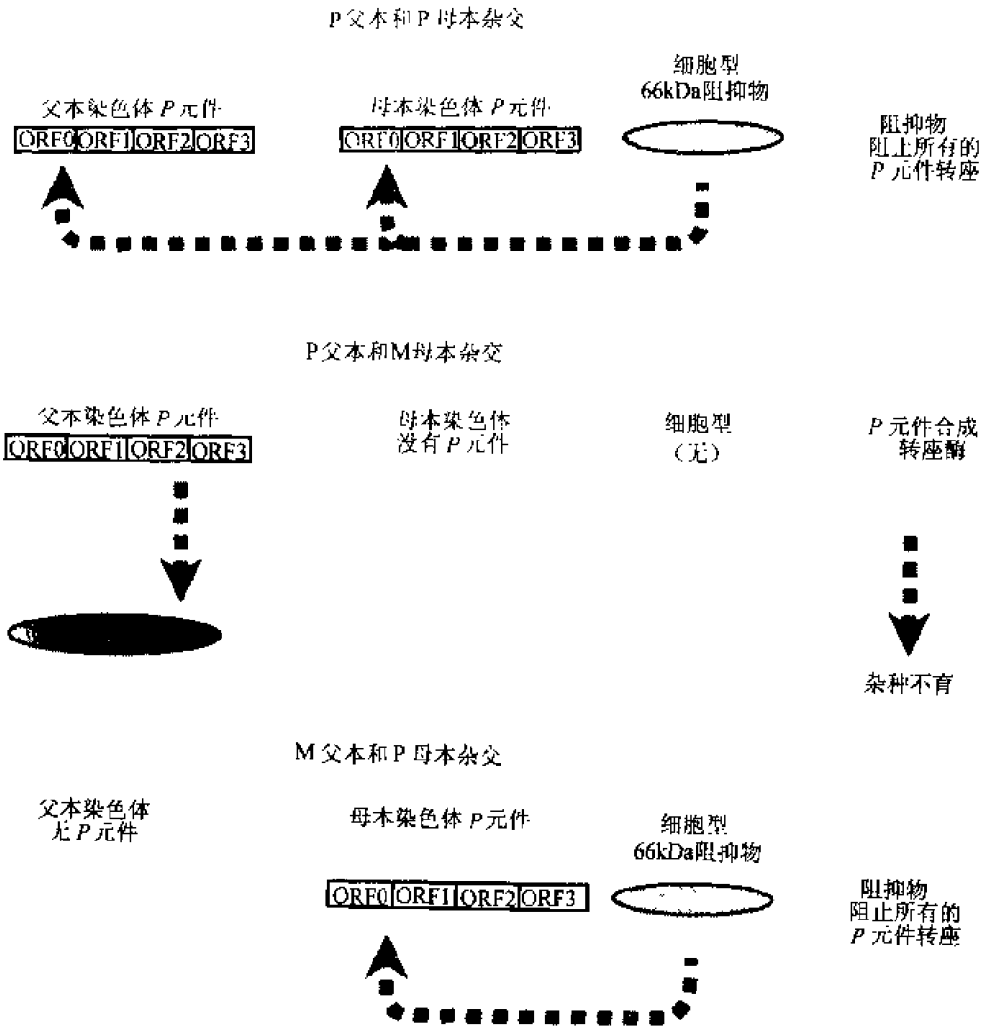


图 10-22 细胞型的影响

杂种不育由基因组中的 *P* 元件和细胞型中 66 kDa 的阻抑物之间的反应所决定。

第六节 反转录病毒和反转座子[11,16,48~61]

一、概 述

转座元件或转座子是基因组中可移动的独立的 DNA 序列，一个转座子可由基因组的一个位置移到另一个位置。从 DNA 到 DNA 的转移过程称转座。从 DNA 到 RNA 再到 DNA 的转移过程叫反转座。后者为经 RNA 介导的转座过程。

必须经过 RNA 中间体的转座是真核生物所特有的，反转录病毒能够将 RNA 病毒基因组的 DNA 拷贝(原病毒)整合到宿主细胞染色体的能力为一典型例证。一些真核细胞的转座子和反转录病毒的原病毒的一般组成结构相关，并且通过 RNA 中间体进行转座。作为一类结构单位被称做反转座子(retroposons，有时亦称 retrotransposons)。它们包含从

反转录病毒自身(能够自由地感染宿主细胞),到自己不能进行转座,但通过 RNA 可以进行转座的序列。它们和其余类型的转座子一样,其共有的可鉴定的特性为:插入位点会产生短的靶序列 DNA 的正向重复。

即使在检测不到活性转座子的基因组中,古老的转座事件所留下的痕迹仍可被发现,其存在形式为散在分布的重复序列两侧的正向重复的靶序列。这些序列的特性有时显示其 DNA 基因组序列的祖先为一个 RNA 序列。我们认为这一 RNA 序列一定先被转化为 DNA 双螺旋拷贝,然后通过与转座相类似的过程插入到基因组。

像任何其他的再生循环一样,反转录病毒或反转座子的再生循环是连续的,因此我们对起始点的选择是随意的。反转录病毒首先被观察到的是以传染性病毒颗粒形式存在,并能在细胞间传播,因此细胞内的循环(包括 DNA 双螺旋)被当作复制 RNA 病毒的方式。反转座子是被当作基因组中的一部分而发现的。RNA 的存在形式在绝大多数情况下根据其功能被当作mRNA。因此我们定义反转座子为基因组的双螺旋 DNA 序列,它们可以在基因组内进行转座,但不在细胞间迁移。

二、反转录病毒和反转座子

1. 反转录病毒的生活周期

反转录病毒具有单链 RNA 基因组,通过双链 DNA 中间体进行复制。病毒的生活周期必须包括一个和转座类似的过程,使双链 DNA 插入到宿主基因组,并使 DNA 靶序列产生短的正向重复序列。

这一反应的重要性超出了仅仅使病毒生存的范围。其结果表现为:

1) 整合到性细胞的反转录病毒序列会作为内源性原病毒(endogenous provirus)而保留在细胞基因组内。和溶原性噬菌体一样,原病毒表现为生物体的遗传物质的一部分。

2) 细胞的序列偶然会与反转录病毒序列相结合并随其发生转座,转座后的序列会以双螺旋形式插入到基因组的新位点。

3) 被反转录病毒所转座的细胞序列,可能使被该病毒感染的细胞改变其特性。

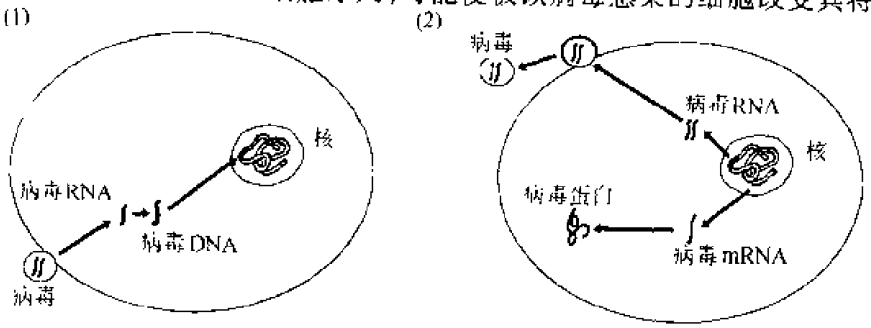


图 10-23 反转录病毒的生活周期

反转录病毒的生活周期 (1)病毒感染宿主细胞,释放出病毒 RNA。在病毒中的反转录酶的作用下,在细胞质中进行反转录,合成线形病毒 DNA(部分可形成环状病毒 DNA)。继而病毒 DNA 整合到宿主细胞染色体 DNA 中,成为原病毒。(2)病毒 DNA 转录生成病毒基因组 RNA;并转录生成病毒 mRNA,进而翻译生成病毒蛋白质,随后包装成子代病毒。

图 10－23 为反转录病毒的生活周期。关键性的步骤为：病毒 RNA 被转换成 DNA，此 DNA 被整合到宿主基因组，随后 DNA 原病毒被转录成 RNA。

负责使 RNA 转化成 DNA 拷贝的酶为反转录酶(reverse transcriptase)，此酶可作为重组 DNA 技术中一个强有力的工具。该酶在被感染的细胞质中使 RNA 转化为线性双螺旋 DNA 分子(这一 DNA 也可被转化为环形，但这种形式似乎并不参与再生循环过程)。

此线性 DNA 进入细胞核。一个或多个 DNA 拷贝被整合到宿主基因组。被称为整合酶(integrase)的酶负责整合反应。被整合的原病毒 DNA 经由宿主的转录系统产生病毒 RNA，此 RNA 既可作为 mRNA，又可作为基因组而包装到病毒粒子中。整合是生活周期的一个正常组成部分，也是转录所需要的。

每个病毒粒子中被包装入两个 RNA 基因组拷贝，因此单一的病毒粒子实际为二倍体。如果细胞被两个不同但却相关的病毒同时感染，则有可能产生携带两个不同 RNA 拷贝的杂合子病毒颗粒。二倍性可能对病毒获取细胞序列是重要的。反转录酶和整合酶随基因组一起被病毒所携带。一个典型的反转录病毒序列含 3 或 4 个“基因”，这一术语在此表示为编码区，每一编码区通过加工实际上产生出多种蛋白质。如图 10－24 所示，一个含 3 个基因的反转录病毒其基因的排列方式为 *gag-pol-env*。

反转录病毒的 mRNA 具有一个通常的结构，在其 5′末端被加帽，并且在 3′末端与多聚腺苷酸相连。它代表两种 mRNA。当全长的 mRNA 被翻译时，所得产物为 Gag 和 Pol 多聚蛋白。当翻译开始于起始密码子而终止于第一个终止密码子时所得产物为 Gag 蛋白。如果要表达 Pol 蛋白，则必须要越过此终止密码。

不同的病毒利用不同的机制可使翻译越过 *gag* 终止码而继续进行，这取决于 *gag* 和 *pol* 可读框的相互关系。当 *gag* 和 *pol* 密码相互连接时，一个可识别终止码的谷氨酰基-tRNA 可以抑制终止过程并合成一个单一蛋白。当 *gag* 和 *pol* 存在于不同的可读框时，一个核糖体的框架迁移可使单一蛋白得到合成。通常情况下，翻译可越过终止码的效率大约为 5%，因此产生的 Gag 蛋白大约为 Gag-Pol 蛋白的 20 倍。

Env 多聚蛋白是通过另一种方式表达的：剪接产生一个短的亚基因组(subgenomic)的 mRNA，并通过它的表达产生 Env 产物。*gag* 基因提供病毒粒子核蛋白核心的蛋白组成部分。*pol* 基因编码参与核酸合成和重组的蛋白。*env* 基因编码参与病毒颗粒外膜成分的蛋白。Gag(或 Gag-Pol)和 Env 蛋白皆为多聚蛋白，它们要经过蛋白酶的切割才能最终形成在成熟病毒颗粒中所发现的单一蛋白质的形式。蛋白酶的活性被病毒以多种形式所编码：它可以是 *gag* 或 *pol* 的一部分，或有时以另一个独立的可读框形式存在。

一个反转录病毒颗粒的形成包括将 RNA 包装于核心部位，在其四周环绕以衣壳蛋白。随后获取了由宿主细胞浆膜与病毒糖蛋白构成的外膜(envelope)形成病毒粒子(virion)并与宿主细胞膜脱离。在感染情况下，这一过程刚好相反：一个病毒颗粒先与细胞浆膜融合，然后释放病毒颗粒的组成部分。

2. 反转录病毒 RNA 转变成病毒线性 DNA 的过程

反转录病毒被称为正链病毒(plus strand virus)，因为病毒 RNA 编码自身蛋白产物。反转录酶负责使基因组(正链 RNA)转化成互补的 DNA 链，此链被称为负链 DNA(minus strand DNA)。反转录酶也催化随后的合成双链 DNA 的反应。它具有 DNA 聚合酶的活

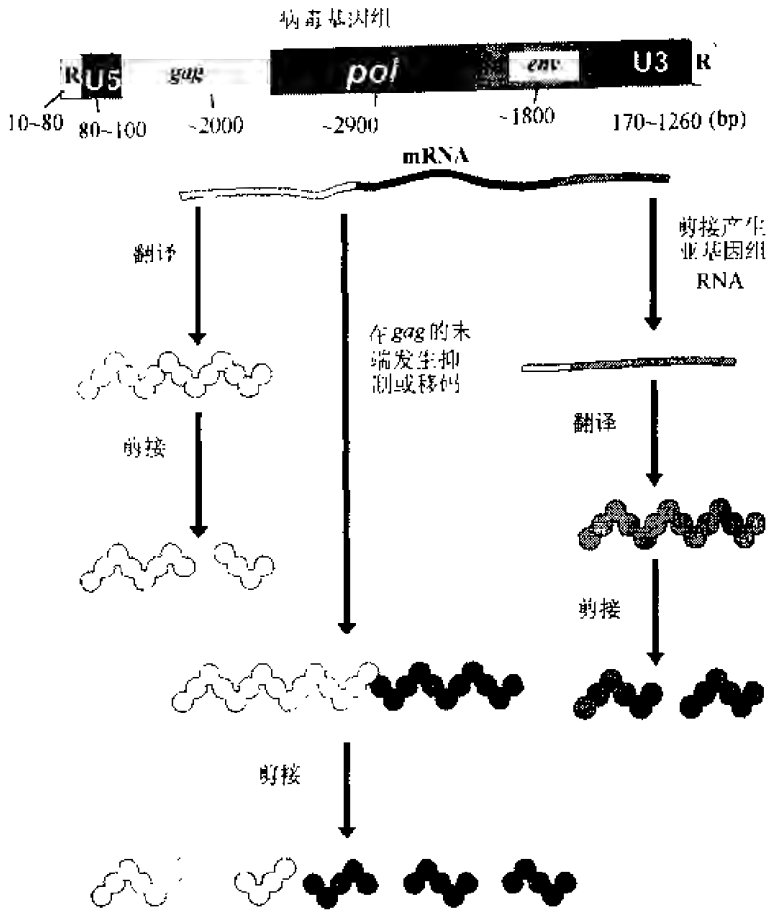


图 10-24　反转录病毒"基因"的表达

反转录病毒的"基因"以多聚蛋白的形式被表达，然后经加工产生独立的蛋白质产物。白圈：Gag，基质(matrix，MA)，衣壳(capsid，CA)和核衣壳(nucleocapsid，NC)。黑圈：Pol，蛋白酶(protease)，反转录酶(RT)和整合酶(IN)；灰圈：表面蛋白(surface protein，SU)和跨膜(transmembrane，TM)。

性，这使其能够由 RNA 的单链反转录合成双螺旋 DNA。此双螺旋的第二条 DNA 链被称为正链 DNA(plus strand DNA)。并且做为活性的必要补充，这个酶具有核糖核酸酶 H(RNase H)的活性，使其能够降解 DNA-RNA 杂合体中的 RNA 部分。所有反转录病毒的反转录酶皆具有相当程度的氨基酸序列相似性，并且在一些别的反转座子中也可发现序列的同源性。

在图 10-25 中，DNA 形式的病毒结构与 RNA 形式进行了比较。在病毒 RNA 的两个末端含有正向重复序列。在不同种病毒中 R 片段有所变化(从 10～80 核苷酸)。在病毒 5′末端 R 片段之后为含 80～100 核苷酸的 U5 区，这一区是 5′端所特有的。在 3′端 R 片段之前为 U3 区(170～1260 nt 长度)，它是 3′端所特有的。当从 RNA 型向 DNA 型转化的过程中，R 片段被用来在线性 DNA 中产生更加长的正向重复序列(见图 10-26)。整合的 DNA 两侧各缺少 2 个碱基对是整合机制作用的结果(见图 10-27)。像其他 DNA 聚合酶一样，反转录酶也需要一个引物。天然的引物为 tRNA，在病毒颗粒中存在一个宿主提供的 tRNA。tRNA 3′端的一个 18 个碱基序列可与病毒 RNA 分子的距 5′端 100～

200 个碱基的序列配对。这个 tRNA 也可以与其它病毒 RNA 的 5′端附近的另一位点相配对，因此有助于病毒 RNA 间二聚体的形成。

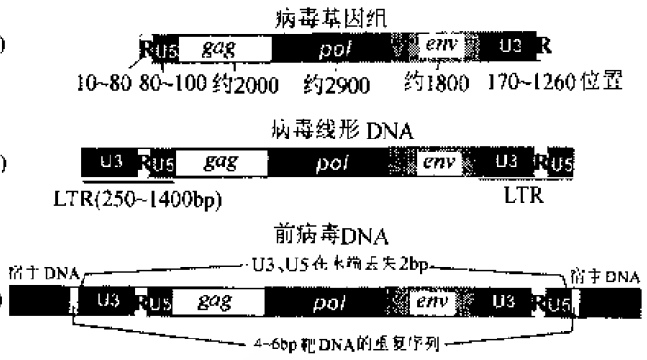


图 10－25　病毒 RNA 形式、DNA 形式和原病毒结构的比较
(1)病毒基因组 RNA 结构；(2)病毒线型双链 DNA 的结构；
(3)整合在宿主细胞染色体 DNA 的原病毒的结构。

这就产生一个难题，反转录酶只能在距 5′端的下游 100～200 碱基处开始 DNA 的合成，如何合成可代表完整 RNA 基因组的 DNA?

在体外，合成进行到末端，产生一个被称为“强停止负链 DNA”(strong stop minus strand DNA)的短 DNA 序列。在体内没有发现此种分子，这是因为反应按图 10－26 所示而继续进行。反转录酶转换了模板，携带这一新合成的 DNA 到一个新的模板。这是二次模板间跳跃的第一次。

在这一反应中，RNA 模板 5′端的 R 区被反转录酶的核糖核酸酶 H 的活性所降解，它的去除使 3′端的 R 区可与新合成的 DNA 相配对。随后反转录继续进行，经过 U3 区而进入到 RNA 的主体。

和“强停止负链 DNA”相配对的 R 区可以来自同一 RNA 分子的 3′端(分子内配对)或不同 RNA 分子的 3′端(分子间配对)。图中所示为分子间配对。

转换和延伸的结果是在 5′端加上了一个 U3 片段。U5－R－U3 这一段序列被称做长末端重复(long terminal repeat，LTR)，因为一个相似的系列反应会在 3′端加上 U5 片段，产生相同的 U5－R－U3 结构。

我们现在需要合成正链 DNA 和另一侧的 LTR。反转录酶开始从原来的 RNA 分子经降解所剩的一个片段来合成正链 DNA。当酶到达模板的末端时则产生一个“强停止正链 DNA”(strong stop plus strand DNA)，这一 DNA 片段随后转移到负链的另一端。它的释放大概是由于从更上游的(图的左侧)引物开始的另一轮 DNA 合成发生后所产生的置换反应。它利用 R 区与 3′端的负链 DNA 相配对。这一双链 DNA 随后需要在两条链完成 DNA 合成，产生一个两侧为 LTR 的 DNA 双螺旋。

因为在每一个反转录病毒颗粒中存在两个 RNA 基因组，因此在病毒的再生循环中有可能发生重组。这有可能发生在两个过程中，一个是负链合成过程中，另一个是正链合成过程中。图 10－26 所示的分子间配对可使两个接连的 RNA 模板的序列发生重组。正链 DNA 的合成可能是不连续的，反应可包括几个内部引发过程。在反应过程中也可以发生链的迁移。

这两种事件的共同特征为重组是 DNA 合成过程中模板交换的结果。这是一种被称为拷贝选择(copy choice)的重组的一个普通例子，有许多年这种重组被认为是一般的重组的一种可能机制。细胞系统利用此机制是不大可能的，但它却是 RNA 病毒感染中重组的一个共同的基础，包括那些完全通过 RNA 形式进行复制的病毒，例如脊髓灰质炎病毒。

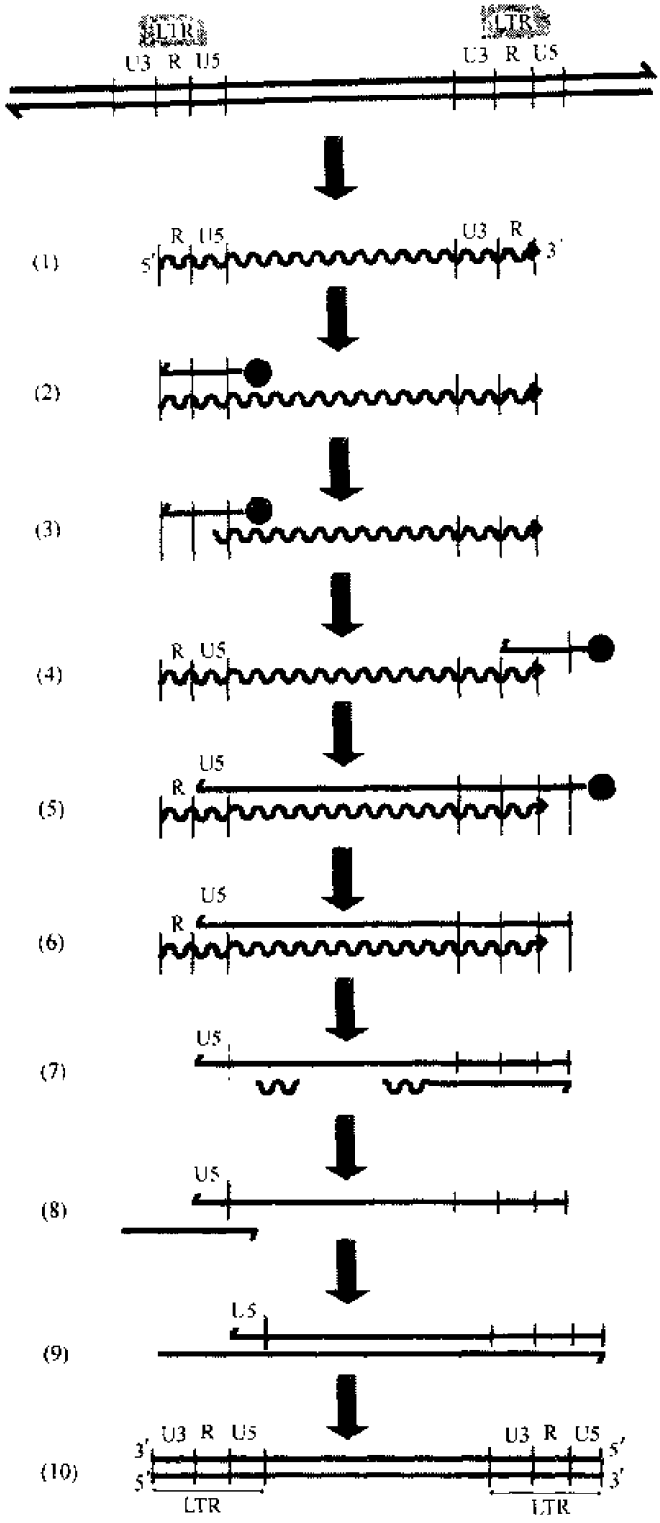


图 10-26　线型病毒 DNA 的合成

(1)反转录病毒正链 RNA;(2)tRNA 引物与反转录病毒 RNA 结合位点退火,反转录酶开始合成负链 DNA,酶到达模板链末端,产生强停止负链 DNA;(3)RNase H 活性降解 RNA 链的 5′末端区;(4)单链 DNA R 区同另一个反转录病毒 RNA 的 3′端配对(完成第一次跳跃);(5)反转录酶继续合成 DNA;(6)tRNA 引物被去除;(7)RNA 被降解,留下的片段作为 DNA 合成的引物,强停止正链 DNA 被合成;(8)正链 DNA 被转移到负链 DNA 的另一个末端(完成第二次跳跃);(9)完成正链 DNA 的合成;(10)完成负链 DNA 的合成。

整合后的原病毒的结构形式类似于线性 DNA，在其两侧为相同的 LTR 序列。在 U5 的 3′末端有一个短的并相对于 U3 的 5′末端的反向重复的序列，因此 LTR 终止于短的反向重复序列。整合的原病毒 DNA 像转座子一样：原病毒序列终止于反向重复序列，并且其两侧为靶 DNA 的短的正向重复序列。

3. 病毒线性 DNA 整合到宿主细胞基因组

原病毒是由线性 DNA 直接插入到靶位点而产生的(除了线性 DNA，病毒序列还以环形存在。一种由于线性末端的连接而含两个相邻的 LTR 序列。另一种只含一个 LTR 序列，这大概是由于重组而形成的并且实际上是环形形式的主体)。虽然有很长时间中间体看起来像是环形的(和 λ DNA 的整合相类似)，我们现在知道线性形式是用来进行整合的。

线性 DNA 在体外的整合已被鉴定，此过程可被单一病毒产物整合酶所催化。整合酶可作用于反转录病毒线性 DNA 和靶 DNA。反应过程如图 10－27 所示。病毒 DNA 的末端是很重要的，正如转座子一样，在末端的突变会阻止整合。其最保守的特征是在接近每一个反向重复序列的末端皆存在一个二核苷酸的序列(CA)。整合酶与 DNA 形成一个核糖核蛋白复合物，从而把 DNA 的两个末端带到一起，并且除去保守的 CA 之外的碱基，把平头末端转化为黏性末端，通常每个 3′末端失去两个碱基。

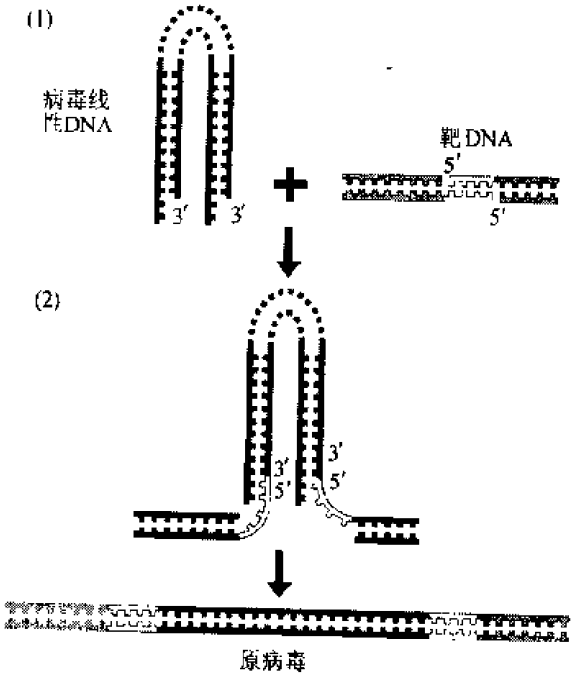


图 10－27　病毒线型 DNA 被整合到基因组
病毒线形 DNA 在整合酶的作用下，整合到宿主细胞的基因组中。在整合中，每个 LTR 失去 2 个 bp。(1)在 LTR 的 3′端，在整合酶的作用下，产生 2 个碱基的黏性末端。整合酶在靶 DNA 产生交错的末端。(2)整合酶连接 LTR 的 3′末端与靶 DNA 的 5′末端。单链区被宿主细胞的酶修复。

靶位点的选择相对于序列来说是随意的。整合酶在靶位点进行交错式切割。在图 10－27 的例子中，切点被 4 个碱基对所分离。靶位点重复序列的长度与病毒的种类有关，它可能是 4、5 或 6 个碱基对。它大既是由整合酶与 DNA 反应的几何特征所决定的。

对靶位点的切割所产生的 5′末端与病毒 DNA 的 3′端发生共价连接。在此点上病毒 DNA 的两个末端皆通过一条链与靶 DNA 相连。单链区被宿主的酶所修复，在这一反应的进程中，病毒 DNA 5′末端的两个突出的碱基被去除。结果经整合的病毒 DNA 的两侧 LTR 皆失去两个碱基。这相当于 5′末端左侧的 U3 区间和 3′末端右侧的 U5 区各失去两个碱基。在整合后的反转录病毒 DNA 的两侧的每个末端存在靶 DNA 特征性的短的正向重复序列。

病毒 DNA 在随意选择的位点整合到宿主基因组中。一个被成功感染的细胞可获得 1～10 个原病毒拷贝(当然，感染性病毒进入的是细胞质，但是它的 DNA 形式却整合到细胞核的基因组中。反转录病毒只能在增殖的细胞中复制，这是因为进入细胞核需要细胞

经过有丝分裂过程，这时病毒基因组能够接近核物质）。

每一个 LTR 的 U3 区含有一个启动子。左侧 LTR 的启动子负责引发原病毒的转录。如前所述，形成原病毒 DNA 时需要把 U3 序列加入到左侧 LTR。因此我们可知启动子是由 RNA 转化成双螺旋 DNA 的过程中产生的。

有时（大概相当少见），右侧 LTR 的启动子负责与整合点相毗邻的宿主序列的转录。LTR 同时携带一个增强子，它可以作用于细胞和病毒序列。一个反转录病毒的整合有可能使某些种类的基因被激活，从而使宿主细胞转变成肿瘤细胞。

被整合的原病毒可以从基因组中被切掉吗？在原病毒的 LTR 序列之间可发生同源性重组。在一些细胞基因组中存在可能是切割反应所遗留的单独的 LTR 序列。

到目前为止已经讨论了反转录病毒的感染循环，说明了整合是进一步产生 RNA 拷贝所需的。但是，当一个病毒 DNA 整合到生殖细胞后，它变成此种生物体的可遗传的“内源性原病毒”。内源性原病毒通常是不被表达的，但有时它们会被外部事件所激活，例如另一种病毒的感染。

4. 反转录病毒的癌基因

病毒生活周期的一个有趣的方面为“转导病毒”（transducing virus）的产生，这些变种已如图 10-28 所示形式获得了细胞的序列。病毒的部分序列被 v-*onc* 基因所代替。蛋白质合成产生一个 Gag-v-Onc 蛋白，而不是通常的 Gag、Pol 和 Env 蛋白。所产生的病毒为复制缺陷型（replication-defective）。其本身不能维持一个感染循环。但是，如有一个能提供所遗失的病毒功能的辅助病毒（helper virus）的伴随可使其生存下去。Onc 为肿瘤生成（oncogenesis，Onc）的缩写，意为能够转化培养细胞使其从通常的生长调控中解脱出来而进行无限制分裂的能力。病毒和细胞的 *onc* 基因都有可能与产生肿瘤细胞有关。

v-*onc* 基因赋予病毒转化某种类型宿主细胞的能力。在宿主基因组中所发现的同源性序列被称为 c-*onc*。反转录病毒是如何获得 *onc* 基因的呢？一个具有启发性的特征为 c-*onc* 和 v-*onc* 基因结构上的差异。c-*onc* 基因通常为内含子所阻断，并且在细胞内低水平表达。v-*onc* 基因是连续的，并且作为病毒转录单位的一部分而高水平表达。这表明 v-*onc* 基因来源于 c-*onc* 基因经剪接的 RNA 拷贝。

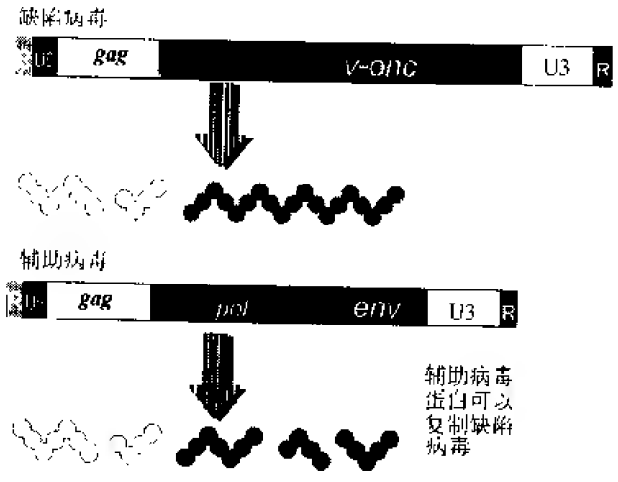


图 10-28 复制缺陷型转化病毒的复制

复制缺陷型转化病毒的部分病毒序列被细胞序列替代。被替代的病毒区和细胞序列的长度对每种病毒是不同的。在具有野生型功能的辅助病毒的协助下，缺陷型病毒可以复制。

转化病毒（transforming virus）的形成机制如下，反转录病毒整合到 c-*onc* 基因附近。缺失过程使原病毒与 c-*onc* 基因融合。随后转录产生一个一端为病毒序列而另一端为 *onc* 序列的融合 RNA。剪接去除此 RNA 中病毒部分和来自细胞部分的内含子。此

RNA 分子含有能使其包装入病毒颗粒的适当信息。如果此细胞中还含有另一个完整的原病毒拷贝，则会产生病毒颗粒。这时一些二倍体病毒颗粒中则有可能含有一个融合 RNA 和一个病毒 RNA。

这些序列之间的重组能够产生其两侧皆为病毒重复序列的转化基因组(transforming genome)。在反转录病毒的感染循环中，各种方式的重组以高频率发生。我们不知其对底物的同源性有何要求，但我们假设病毒基因组与融合 RNA 的来自细胞部分之间的非同源性的重组机制和负责病毒重组的反应机制相同。

反转录病毒类的共同特性表明它们有可能来源于一个祖先。原始的插入序列 IS 元件可能先包绕一个宿主核酸聚合酶基因；其结果为 *LTR-pol-LTR* 单位形成。它有可能通过获得更加复杂的操作 DNA 和 RNA 底物的能力(包括掺入其产物负责包装 RNA 的基因)而进化成感染性病毒。其他的功能，例如转化基因(transformig gene)，可能在较后期被掺入(获取细胞功能的机制不是 *onc* 基因所特有的，但是携带这些基因有可能提供一个选择性优势，因为它可刺激细胞生长)。

三、反转座子的分类

反转录病毒为反转座子的范例，它们的转座能力是由于其编码反转录酶和/或整合酶。反转座子与反转录病毒的差别在于其不经过一个独立的具感染性的形式，但其转座的机制相类似。这一类被称为病毒类超家族(viral superfamily)。

另一类反转座子的内部的和外部的特性表明它们来源于 RNA 序列，虽然在这种情况下我们只能推测 DNA 的拷贝是如何产生的。我们假定它们是由别的位置所编码的负责转座的酶系统的底物。它们来源于细胞内的转录物。它们并不编码具有转座功能的蛋白质，这一类被称为非病毒类超家族(nonviral superfamily)。

哺乳动物的基因组含大量的相对来说较短的彼此相关的序列(包括中等程度的重复性 DNA 序列)。反转座子是其中一个主要的组成部分，这种元件的绝大部分属于两个家族，它们最初被鉴别为散在分布的重复序列，每一家族皆包含在基因组中散在分布的许多成员，*LINE*(long interspersed repeated sequences)包含散在分布的长重复序列，*SINE*(short interspersed repeated sequences)包含散在分布的短重复序列。另一个重要的区别也许在于 *LINE* 来源于 RNA 聚合酶 II 的转录物，而 *SINE* 来源于 RNA 聚合酶 III 的转录物。

(一) 病毒类反转座子

1. 酵母 Ty 元件

Ty 元件为一族在不同酵母株的不同位点上散在分布的 DNA 重复序列。Ty 是酵母转座子(transposon yeast，Ty)的缩写。一个转座事件产生一个特殊痕迹：在被插入的 Ty 元件两侧靶 DNA 各有 5 个碱基对长的重复序列。Ty 转座子的转座频率要比细菌转座子的低，大约 $10^{-7}\sim10^{-8}$。

表 10-1 反转座子可分为病毒和非病毒两类超家族

	病毒类	非病毒类
一般类型	反转录病毒类 Ty(酵母) *copia*(果蝇) *LINE L1*(哺乳类)	*SINE B1 /Alu*(哺乳类) 聚合酶 III 的转录物 被加工的假基因
末端	长末端重复序列	一般无重复序列
靶 DNA 重复序列	4~6 bp	7~21 bp
可读框	反转录酶和/或整合酶	无(或不编码转座子产物)
组织结构	可能含内含子 (在亚基因组 mRNA 中被删除)	无内含子

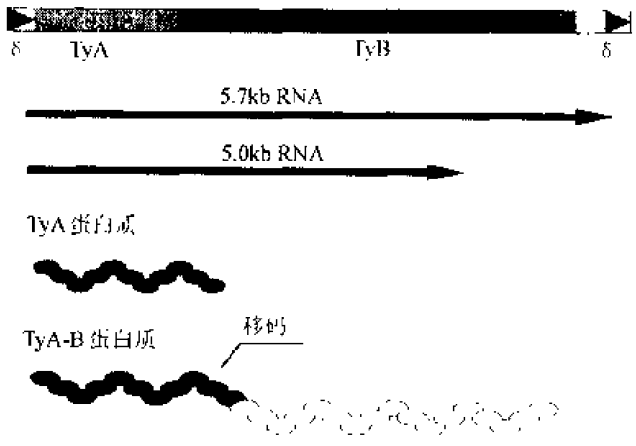


图 10-29 Ty 元件的结构和转录
Ty 元件末端为短的正向重复序列并被转录成重叠的 RNA。它们有两个可读框并与反转录病毒的 *gag* 和 *pol* 基因有关。

在每个 Ty 元件之间存在相当的差异性。大多数元件属于称为 Ty1 和 Ty 917 的两个主要类别。如图 10-29 所示,它们含有相同的一般组织结构,每一结构单位长 6.3 kb,在每一末端含被称为 δ 的 330 bp 正向重复序列。每一类型中的每个 Ty 元件皆与此类的原型结构有很多差异,这包括碱基对的取代、插入和缺失。在一个典型的酵母基因组中存在大约 30 个 Ty1 类的拷贝和大约 6 个 Ty 917 类的拷贝,除此之外,还含有大约 100 个独立的 δ 结构(被称为 δ 序列,solo δs)。虽然在一个 Ty 元件两侧的重复序列很可能是相同的或至少是非常相近的,δ 序列也显示出相当的结构多样性。和 Ty 元件相连的 δ 序列比独立的 δ 序列显示出更高的保守性,这表明在转座过程中涉及对其重复序列的识别。

Ty 元件被转录成两种 poly(A)$^+$ RNA 序列,其组成可占单倍体酵母细胞总 mRNA 的 5% 以上。它们都是起始于左侧 δ 结构的启动子。一个终止于 5 kb 之后处,另一个终止于 5.7 kb 之后处(位于右侧的 δ 序列内)。

Ty 元件的序列含有两个可读框,表达的方向相同,但不在同一阶段阅读并且含 13 个氨基酸残基的重叠。TyA 的序列编码一种 DNA 结合蛋白,TyB 包含与反转录病毒的反转录酶、蛋白酶和整合酶序列同源的序列。

TyA 和 TyB 的结构与功能与反转录病毒的 *gag* 和 *pol* 相似。TyA 和 TyB 的可读框以两种方式表达。Tya 蛋白代表 TyA 可读框,并且在其末端终止。但是,TyB 可读框架只以连接蛋白的一部分被表达,在这里 TyA 和 TyB 区被一个特殊的框架迁移所连接,这使终止码被绕过(与反转录病毒的 *gag-pol* 翻译相似)。

Ty 元件提供了可迁移的同源性序列区，这些区域可成为由宿主细胞所控制的同源性重组的目标。这一过程的通常倾向是对染色体的损伤，它可以造成序列的缺失，或倒位（通过一个染色体上的两个 Ty 元件间的重组），或当重组发生在两个位于不同染色体上的 Ty 元件时，造成更显著的变化。

Ty 元件间重组的发生似乎是爆发式的。当一个重组事件被检测到后，发现别的重组的可能性增加。位于不同位点的 Ty 元件可发生基因转换，其结果是一个结构序列被另一个所取代。

正向重复序列 δ 序列之间的同源性重组可以使 Ty 元件单位被去除。大量独立 δ 序列的存在有可能是此种事件留下的痕迹。由于插入 Ty 元件而产生的突变有可能通过此种方式而使序列恢复原状；恢复的水平则取决于所留下的 δ 序列的精确结构。

值得注意之处在于两个 δ 结构的序列相同，但是在一端的 δ 序列有一个活性的启动子，而另一端则含一个终止密码（别的可转座的结构，包括反转录病毒，含有相似的结构特征）。

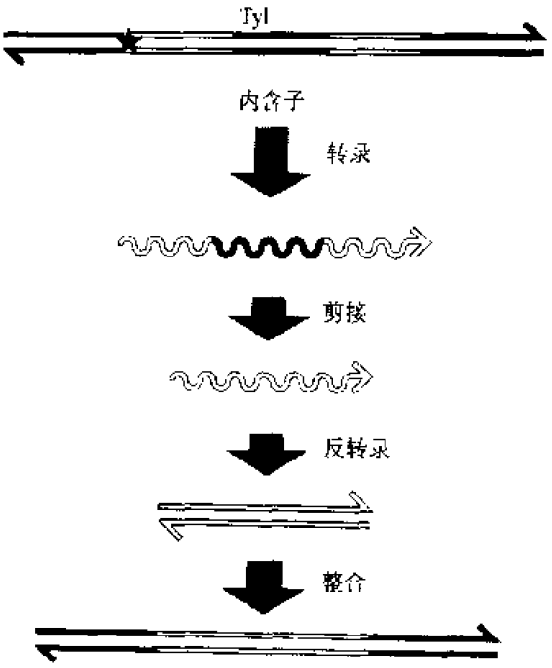


图 10－30　一种特殊的 Ty 元件
一个特殊的 Ty 元件，人工加入一个内含子，转座后产生一个缺乏内含子的拷贝，拷贝具有同样的末端重复序列，产生于原来的 Ty 元件的侧末端。转座的频率与控制转录的启动子有关。

Ty 元件是典型的反转座子，其转座需通过 RNA 中间体。图 10－30 所示为一个巧妙的检测方法。在一个 Ty 元件中插入一个内含子使其成为一个独特的结构，这一序列被置于质粒中，并且使其受 *GAL* 启动子的控制，然后质粒被引入酵母细胞中。转座的结果为在酵母基因组中出现此转座子的许多拷贝，但它们皆不含内含子。

我们知道只有一种去除内含子的方法：RNA 剪接。这表明其转座与反转录病毒的转座机制是相同的。Ty 元件先被转录成 RNA，然后由剪接系统所识别。被剪接的 RNA 为反转录酶所识别并产生双螺旋 DNA 拷贝。

与反转录病毒还有更进一步的相似性，初始的 Ty 元件中的两个 δ 序列存在差异性。但是转座后的结构含相同的 δ 序列，其来源于初始的 5′δ 序列。如果我们考虑 δ 序列和 LTR 一样含有 U3－R－U5 区，Ty RNA 为从一端的 R 区延展到另一端的 R 区，如图 10－24～图 10－26 反转录病毒所示的那样，完整的 LTR 由把 U5 加到 3′端，U3 加到 5′端的操作所形成。

转座是由位于 Ty 元件中的基因所控制。用来控制被标识的 Ty 元件转录的 *GAL* 启动子是可被诱导的：它可由加入半乳糖而启动。启动子的诱导有两种效果，它是激活被标识的元件转座所需要的，并且它的激活会增加酵母染色体 Ty 元件的转座频率。这表明 Ty 元件的产物可以反式作用于其他的元件。

虽然 Ty 元件并不产生感染性颗粒，但病毒样颗粒(virus-like particle，VLP)在已经诱导转座的细胞内堆积，这些颗粒包含全长 RNA、双链 DNA、反转录酶活性和具有整合酶活性的 TyB 产物、TyA 的产物，像 *gag* 的前体一样被切割产生 VLP 的成熟的核心蛋白。这使 Ty 转座子和反转录病毒的类似性更进了一步。简而言之，Ty 元件表现得就像反转录病毒失去 *env* 基因而因此不能恰当地包装它的基因组一样。

在所有酵母基因组中，只有一些 Ty 元件具有活性，绝大多数则失去了转座的能力(相似于惰性的内源性原病毒)。但是，因为这些"死亡"的元件含 δ 重复序列，在活性 Ty 元件合成的蛋白质作用下它们提供了转座的靶体。

2. 果蝇可转座的元件

在果蝇中存在可转座的元件的推断，来源于与大肠杆菌插入序列相似的实验观察。研究发现不稳定的突变可以通过缺失而转化成野生型，如图 10-31 所示，它们是由几类可转座的序列所组成的，其中包括 *copia* 反转座子、*FB* 家族和 *P* 元件。

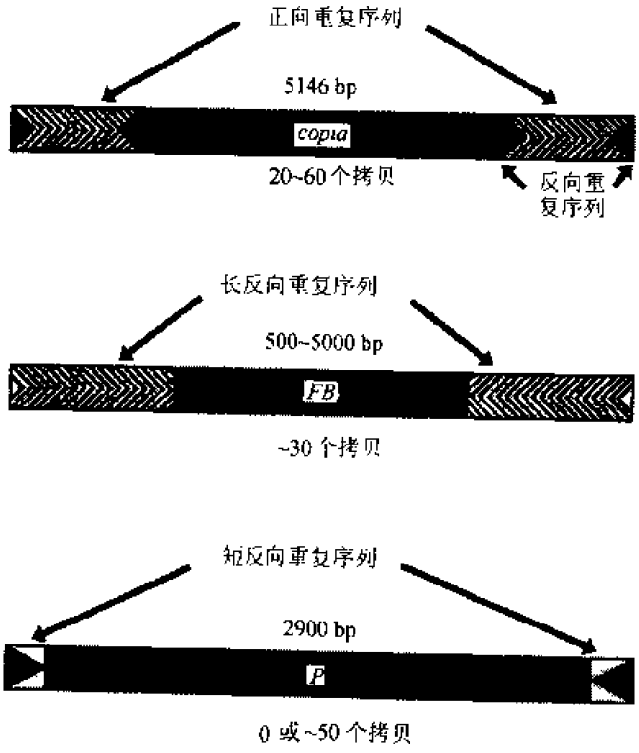


图 10-31 果蝇三种不同类型的转座元件

被充分定性的反转座子家族是一类被称为 *copia* 的反转座子。它的名称表明编码高丰度 mRNA 的大量的密切相关联序列的存在。*copia* 元件的拷贝数取决于果蝇品系的类型，通常为 20～60 个。此类元件散在分布。每一品系果蝇的 *copia* 元件的分布谱虽然有重叠，但并不完全相同。

这些差别是通过进化累积而形成的。对近 40 年左右在实验室中由于繁殖而产生的分化品系间的比较几乎没有显示什么变化。我们不能估计变化的速率，但培养的细胞的变化结果则显示了这种过程的特性。每个基因组中 *copia* 元件的个数可发生显著增加，可达 2～3 倍。增加的元件表明在新的位点有 *copia* 的插入。培养器中某种未知因素的变化可以在短时间内使转座的发生率提高到每代 10^{-3}～10^{-4}。*copia* 元件大约 5000 bp 长，其两侧为 276 bp 的完全一样的正向重复序列，每一个正向重复序列的两侧为相关的反向重复序列。在插入位点会产生靶 DNA 的 5 bp 长正向重复序列。*copia* 家族的各成员之间只有很小的趋异性(divergence)，通常小于 5%。差异通常包含少量的缺失。所有这些特性都是其他 *copia* 样家族所共有的，虽然它们各成员之间显示出更高的趋异性。

每个 *copia* 元件两侧的正向重复序列完全相同，表明它们之间的相互作用允许修正

反应的发生，或者在转座过程中它们皆产生于先代的正向重复序列之一。Ty 元件的情况与其相类似，这对阐明同反转录病毒之间的关系是很有启发的。

在基因组中 *copia* 元件总是完整的。尚未检测到单独存在的末端重复序列。*copia* 元件有时以游离的环型 DNA 的形式存在。与反转录病毒环形 DNA 相类似，较长的一种含两个末端重复序列，较短的一种只有一个。曾经发现含 *copia* RNA 的颗粒。

copia 序列含一个 4227 bp 长的单一可读框。*copia* 可读框的某些部分与反转录病毒的 *gag* 和 *pol* 序列具有同源性。一个值得注意的特性为 *copia* 序列与负责反转录病毒包膜的反转录病毒 *env* 序列不具同源性，这表明 *copia* 很可能不能产生与病毒相类似的颗粒结构。

copia 转录的产物以大量 poly(A)$^+$ mRNA 的形式存在，其含有全长的或部分的转录物。这些 mRNA 含有共同的 5′末端，这是因为转录引发于一个末端重复序列的中部。几种蛋白质由此而产生，它们大概与 RNA 的剪接和多聚蛋白质的剪接过程有关。

虽然我们缺乏有关 *copia* 转座方式的直接证据。但由于其与反转录病毒的组织结构具有如此多的相似性，我们会得出如下结论：*copia* 的起源肯定与反转录病毒相关。很难说 *copia* 具有多少反转录病毒的功能。当然，我们知道它可以转座，但和 Ty 元件类似，没有证据表明它具有任何传染的能力。

在果蝇中含另一族被称为 *FB*(fold back，FB)的反转座子，其成员含不同长度的反向末端重复序列。一些 *FB* 元件仅含有并排的反向重复序列，另一些则在反向重复序列之间含一个非重复性的 DNA 区。

尽管长度不同，*FB* 族所有成员的反向重复序列皆具同源性。其结构解释了这一特性：反向重复序列包含串联式的简单序列 DNA 拷贝，被较长的更具趋异性的序列所分隔。从末端进入到元件内部，简单序列的长度增加，开始其长度为 10 bp，随后扩展到 20 bp，最终扩展到 31 bp。*FB* 元件两侧的两个反向重复序列并不相同。

FB 元件末端的结构产生了一个问题。我们不知道是什么赋予了它转座的能力。有时两个不完全相同的 *FB* 元件很显然以协同方式使介于其间的一大段 DNA 片段发生转座，其方式大概与复合型细菌转座子相似。(虽然介于 *FB* 元件间的 DNA 片段可以长许多，可达 200 kb)。有可能任何两边具有 *FB* 末端的 DNA 序列，皆可作为这种结构单位。

3. 来源于 RNA 聚合酶 II 转录产物的反转座子

哺乳动物基因组包含 20～50 000 个被称为 *L1* 的 *LINE* 序列。其典型的一员约 6 500 bp 长，末端富含腺嘌呤核苷酸，并且有可能存在可读框。一个被测序的元件含两个有 14 bp 重叠的各长 1137 bp 和 3900 bp 的可读框。可以发现转录物的存在。由于重复性 DNA 的存在，*LINE* 家族各成员之间存在差异性。但和种间的差异性相比较，同一种内的各成员之间显示相对较高的同源性。

图 10-32 对病毒类超家族的各成员与反转录病毒进行了比较。我们知道一个具有活性的 Ty 元件编码转座功能。我们也许可以推断在果蝇基因组的 *copia* 序列之中一定含有一些活性元件来编码转座功能。

LINE 最初被划为非病毒一类。但序列分析显示 *L1* 的可读框与反转录酶具有同源

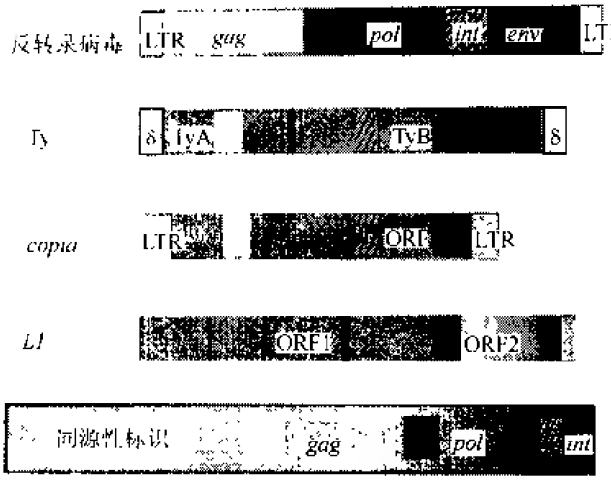


图 10－32　病毒类家族反转座子的结构
病毒类家族的反转座子有末端重复序列和可读框

性。这表明 *L1* 有可能来源于一个编码自身转座的可移动基因，因此可以把它当作病毒类超家族的一员。但我们尚不知道 *LINE* 元件是否编码功能性蛋白或只是由于其起源的结果而与反转座子保持同源性。

LINE 元件和一些其他的元件，并不像典型的反转录病毒那样以 LTR 为末端结构。这就产生一个问题：反转录是如何引发的？因为它不包含引物 tRNA 与 LTR 的配对典型反应（见图 10－26）。在这些元件的可读框中缺乏许多反转录病毒的功能，例如蛋白酶或整合酶区域结构，但在典型情况下确实含有与反转录病毒相类似的序列，并且其产物也许还含有核酸内切酶的活性。

图 10－33 显示这些功能如何支持转座反应。首先由反转座子所编码的核酸内切酶在靶 DNA 上制造一个切口。反转座子的 RNA 产物与结合在切口处的蛋白相缔合。切口可提供 3′－OH 末端用来引发以 RNA 为模板的 cDNA 的合成，为了打开另一个 DNA 链，需要第二次切割过程，随后杂化 RNA/DNA 链或者在此时，或者在转化成 DNA 双螺旋之后与切口的另一端相结合。某些可移动内含子的移动则采用与此相似的机制。

因为 LINE 来源于 RNA 聚合酶 II 的转录物，其基因组中的序列必然是无活性的。它们缺少位于转录起始点上游处的启动子。因为它们通常具有成熟转录物的特性，它们被称为加工过的假基因（processed pseudogenes）。

原则上任意一个 RNA 聚合酶 II 的转录物皆可产生此种假基因，并且有许多这种例子，包括首先发现的剪接过的珠蛋白假基因。

假基因的起始点相当于 RNA 的 5′末端，只有在 DNA 来源于 RNA 的情况下，此种情况才可发生。几个假基因包含精确相连的外显子序列，我们知道没有什么机制可以识别位于 DNA 中的内含子，这种特性表明可能存在一个 RNA 中间体的阶段。假基因可以终止于一小段 A-T 碱基对，这有可能来源于 RNA 的 poly(A)末端结构。在假基因的两侧各含一小段正向重复序列，这有可能产生于和转座相类似的过程。剪接过的假基因位于与它们假定的初始位点不相关的位点上。

剪接过的假基因不含任何可能被用来引发转座过程（或对 RNA 进行反转录）的信息。这一过程是否可能是由反转录病毒所介导的？或由一个异常的细胞系统所完成的？也许转座的序列的末端与转座子的末端的相似性仅仅是偶然的。

转座过程目前正在这些基因组中发生还是我们所见到的仅仅是古老的系统所留下的痕迹？注意：为了使转座得以生存，它们必须在性细胞中发生过，假设在体细胞中也有相类似过程发生，但它们的存在不超过一代。

（二）非病毒类反转座子

最重要的 *SINE* 是由属于单一家族的各成员组成。其短的序列和高度的重复性使其与简单序列 DNA 相类似，不同之处在于家族中的每个成员在基因组中散在分布，而不是局限于串联簇中。与种间的差异性相比较，一个种内各成员之间同样会有显著的相似性。

在人的基因组中，大部分中等程度重复性 DNA 序列约为 300 bp 长的 DNA 序列，散在地分布于非重复序列之间。至少一半的双螺旋 DNA 可被限制性内切酶 AluI 在位于序列 170bp 的单一位点处切断。

被切断的序列皆属于同一家族的成员，这一家族由于鉴定的方式而被称为 *Alu* 家族。在单倍体的基因组中其数量大约为 300 000 个。（相当于 6 kb 长的 DNA 中含一个此序列）。单一的 *Alu* 序列十分广泛的散在分布。相关的序列家族存在于鼠（被称为 B1 簇，含 50 000 个成员）、中国仓鼠（被称为 *Alu* 等同家族）和其他一些哺乳动物中。

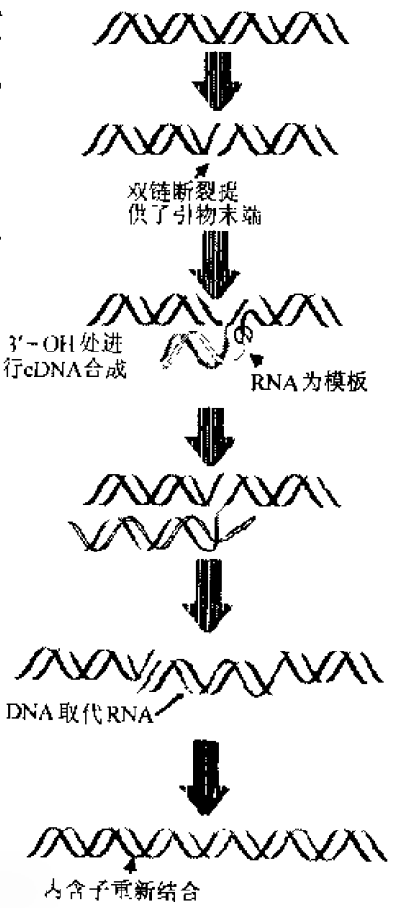


图 10-33 非 LTR 元件的反转座
（1）双链断裂提供引物末端；（2）以 RNA 为模板由 3′-OH 合成 cDNA；（3）DNA 置换 RNA；（4）重组整合。

Alu 家族中各成员之间是相关的但并不完全一样。人类一家族似乎来源于 130 bp 的串联式复制，并且有一个 31 bp 长的不相关序列插入于二聚体的右一半。这两个重复序列有时被称为 *Alu* 序列的左一半和右一半。*Alu* 家族每个成员与原型序列的平均一致性为 87%。鼠的 B1 重复单位的长度为 130 bp，相当于人的一个单位二聚体中的一个单体。它与人的序列的同源性为 70%～80%。

Alu 序列与 7SL RNA 相关联，7SL 为信号识别单位的一个组成部分。7SL 与 *Alu* 序列的左一半相对应，并且中间含一个插入序列。因此 7SL RNA 的 5′末端的 90 个碱基与 *Alu* 的左侧同源，7SL RNA 中间的 160 个碱基与 *Alu* 没有同源性，7SL RNA 3′末端的 40 个碱基与 *Alu* 的右侧同源。编码 7SL RNA 的基因被 RNA 聚合酶 III 活跃地转录。有可能非活性的 *Alu* 序列产生于这些基因（或其相关的基因）。

Alu 家族各成员与转座子的相似性在于其两侧为短的正向重复序列。但是，奇怪的是家族中各成员之间的这种重复序列的长度并不相同。因为它们来源于 RNA 聚合酶 III 的转录物，有可能每个成员含有内部的活性启动子。

已知 *Alu* 家族具有各种各样的性质，它的普遍存在促使了许多关于其功能的看法的提出，但目前尚不能确定其真正的作用。

Alu 序列含有一个 14 bp 区域，它与位于乳多空病毒（papova virus），例如 SV40 和乙

型肝炎病毒复制起始点的一段序列几乎完全一致。这就产生一种可能性：*Alu* 家族可能与真核生物基因组的复制起始位点相关联，虽然，*Alu* 家族成员的数量与此种推论不一致，*Alu* 序列的总数大概是预料的复制起始点的十倍。

至少 *Alu* 家族中的一些成员可以被转录成独立的RNA。在中国仓鼠中，*Alu* 家族等同族的一些成员（虽然不是全部）显示其可以在体内被转录。这类转录单位被发现位于其他转录单位的附近。

Alu 家族的成员有可能被包括于结构基因转录单位中，例如其存在于长的核 RNA 中。在单一核分子中多个 *Alu* 拷贝的存在可以产生二级结构。事实上，*Alu* 家族的成员以反向重复序列的方式的存在是在哺乳动物核 RNA 中，为多数二级结构的形成原因。这一类反转座子具有经 RNA 转座的特征，因为它们都缺乏与反转座有关的基因，因此这类反转座子的返转座可能是被动的，其确切机制目前还不清楚。

（张玉静　阮承迈）

参 考 文 献

[1] Lewin B. Genes. Oxford University Press, 1997 563
[2] Davis R F L. Genetic Recombination. Blackwell Sciences Ltd, 1996
[3] Kleckner N. Cell, 1977 11:11
[4] Calos M et al. Cell, 1980. 20:579
[5] Kleckner N. Ann Rev Genet, 1981. 15:341
[6] Campbell A. Ann Rev Microbiol, 1981 35:55
[7] Brock D J. Molecular Genetics for the Clinicm. Cambridge: Cambridge University Press, 1993
[8] Kaiser K et al. Mobile Genetic Elements: Frontiers in Molecular Biology. Oxford: Oxford University Press, 1995. 690
[9] Craigie R. Cell, 1985. 41:867
[10] Craigie R. Cell, 1986. 45:793
[11] Engelman A et al. Cell, 1991. 67:1211
[12] Mizuuchi K. Cell, 1991. 66:129
[13] Mizuuchi K. Cel., 1983. 35:785
[14] Mizuchi K. Cell, 1984. 39:395
[15] Sherratt D(ed). Mobile Genetic Elements: Frontiers in Molecular Biology Oxford: Oxford University Press, 1995
[16] Van Luenen H G et al. Cell, 1994. 79:293
[17] Francia M V et al. J Bacteriol, 1996. 178:894
[18] Golden J W et al. Nature, 1985 327:419
[19] Kanaar R et al. Proc Natl Acad Sci USA, 1988. 85:752
[20] Kunkel B et al. Genes Dev, 1990. 4:525
[21] May E W et al. Science, 1996. 272:401
[22] Johnsrud L et al. Cell, 1978. 15:1209
[23] Van De Putte et al. J Bacteriol, 1984. 158:517
[24] Craig N L. Cellular and Molecular Biology, 2nd ed. Washington, D.C.: ASM Press, 1996
[25] Aldaz H et al. Cell, 1996. 85:257
[26] Savilahti H et al. Cell, 1996. 85:271
[27] Droge P et al. Proc Natl Acad Sci USA, 1990. 87:5336
[28] Kleckner N. Ann Rev Cell Biol, 1990 6:297
[29] Mizuuchi K. Ann Rev Biochem, 1992 61:1011
[30] Reznikoff W S. Ann Rev Microbiol, 1993. 47:945
[31] Scott J R. Ann Rev Microbiol, 1995. 49:367

[32] Shapiro J A. Proc Natl Acad Sci USA, 1979. 76:1933
[33] Hanford D B et al. Cell, 1981. 64:171
[34] Roberts D et al. Cell, 1985. 43:117
[35] Grindley N D. Cell, 1978. 13:419
[36] Bolland S et al. Cell, 1996 84:223
[37] Bender J et al. Cell, 1986. 45:801
[38] Bissonnette L et al. J Bacteriol, 1992. 174:1248
[39] Foster T J et al. Cell, 1981. 23:201
[40] Grindley N D et al. Cell, 1982. 30:19
[41] Gierl A et al. Ann Rev Genet, 1989. 23:71
[42] Simon M et al. Science, 209:1370
[43] Engels W R. Ann Rev Genet, 1983. 17:315
[44] Grindley N D et al. Ann Rev Biochem, 1985. 54:863
[45] Berg D E et al. Mobile DNA. Washington D. C.: American Society of Microbiology, 1989
[46] Laski F A et al. Cell, 1986. 44:7
[47] Spradling A C et al. Science, 1982. 218:341
[48] Finnegan D. J Int Rev Cytol, 1985. 93:281
[49] Varmus H et al. Mobile DNA, Eds M M Howe and D E Berg. Washington D. C.: American Society for Microbiology, 1989. 53
[50] Sherrat D. Mobile Genetic Elements: Frontiers in Molecular Bialogy. Oxford: Oxford University Press, 1995
[51] Craigie R et al. Cell, 1990. 67:829
[52] Hu W S et al. Science, 1990. 250:1227
[53] Goff S P. Ann Rev Genet, 1992. 26:527
[54] Lai M M. Microbiol Rev, 1992. 56:61
[55] Katz R A et al. Ann Rev Biochem, 1994. 63:133
[56] Boeke J D et al. Cell, 1985. 40:491
[57] Mount S M et al. Mol Cell Biol, 1985. 5:1630
[58] Hutchison C A et al. Mobile DNA, Eds M M Howe and D E Berg. Washington D. C.: American society for Microbiology, 1989. 593
[59] Luan D D et al. Cell, 1993. 72:595
[60] Weiner A M et al. Ann Rev Biochem, 1986. 55:631
[61] Deininger P L. Mobile DNA, Eds M M Howe and D E Berg. Washington D. C.: American Society for Microbiology, 1989. 619

第十一章　DNA 的重排和扩增

第一节　概　　述[1~2]

在一个物种各个成员之间，基因组的一般结构是恒定的。然而，在某些情况下，在一个基因组内，一些 DNA 序列可以被移动，被修饰，被扩增，甚至丢失。新的序列通过实验手段被引入细胞时，它们也可以被整合到基因组内。基因组的可变性允许我们对基因组进行改造。

在低等真核生物中，特异序列的重排或丢失的例子是很常见的，通常这些改变只涉及到体细胞，性细胞则保持不变（然而，在有些生物体的生殖周期中，涉及到整个染色体或染色体组的丢失）。虽然免疫系统存在着特殊序列的大量的重排，但在动物体内，一个序列的重新组织是很罕见的。

在真核和原核生物中，重排可产生两种结果：

1）重排可以产生在特殊情况下需要表达的新基因，免疫球蛋白就是一个例子。

2）重排可以改变基因的表达，由表达以前存在的某一个基因转换为另一个基因，它提供了一个调节基因表达的机制。

酵母接合型的转变和锥虫抗原的变异经历相似的途径，在该过程中，基因的表达由对 DNA 序列的操作所控制。表型由存在于一个特殊的、活性基因座的基因拷贝所决定。但基因组也贮存有另一些不表达的序列。沉默的基因拷贝只能通过 DNA 序列的重排而被活化，在此重排过程中它替代了活性基因拷贝，这种替换等同于一个特定靶部位的单向转座。

这种途径最简单的例子见于酿酒酵母（*S. cerevisiae*）。单倍体酿酒酵母拥有两种接合型中的任意一种，接合型由存在于活性接合型基因座处的序列所决定。但是，基因组也含有另外两个沉默基因座，一个代表一种接合型。接合型的转换是通过特定序列的替换而实现的，即由携带另一种接合型的沉默基因序列替代活性基因的序列。

非洲锥虫（African trypanosomes），一种单细胞寄生虫，能通过改变其表面抗原而逃逸宿主的免疫反应，此类变异是通过 DNA 重排而实现的。表面抗原的类型是由位于一个活性基因座中的基因序列所决定的。这个序列是可以改变的，它可以被许多沉默基因中的任意一个序列所替代。它似乎是利用与免疫器官产生多样性相似的机制来对抗免疫系统，即依赖于基因组中 DNA 的重排来改变表达序列。

另一种增加遗传容量的方式见于寄生虫或共生物与寄主之间的相互作用，在相互作用中外源性 DNA 由细菌引入到宿主细胞。其机制类似于细菌的接合。外源性 DNA 在新的宿主细胞内的表达改变了细胞的表型。在根癌农杆菌（*Agrobacterium tumefaciens*）的例子中，因为它在被感染的植物细胞中诱导肿瘤的形成，因此其作用是有害的。

在体细胞发育时，基因组成分相对比例的改变，可使昆虫幼虫的某些基因的拷贝数增加。在培养的哺乳动物细胞中有时有基因扩增，因此，我们能选择某些基因拷贝数增加的变异细胞。在基因组内，扩增过程产生的附加的基因拷贝可以染色体内或染色体外的形式存在。

当外源性 DNA 被引入到真核细胞时，它可以染色体外的形式存在或被整合到基因组中，染色体外形式或基因组的形式之间的关系是无规律的，取决于机会和在某种程度上不能预料的事件，而不像细菌质粒的游离型和整合型那样有规律地进行互换。

然而，不管是怎样完成的，该过程可导致基因组发生稳定的变化。当 DNA 被注入到卵细胞之后，DNA 甚至可以被整合到基因组并作为一个正常成分被遗传，有时还能维持其功能。被注入的 DNA 可以进入性细胞和体细胞，产生一个转基因动物。引入特异的并且能以适当的形式发挥其功能的基因的能力，可能成为治疗遗传病的一种主要的医学手段。

与引入新基因相反的手段是能够干扰内源性特异基因的能力。一个额外的 DNA 片段可被引入到一个基因内并阻止该基因的表达，产生一个无效等位基因。饲养携带一个无效的等位基因的动物，能生育一个没有该基因活性拷贝的纯合子动物。这是研究一个基因的功能和重要性的一种直接和有效的方法。

因此，大量的 DNA 序列的实验操作是通过权威性的科学设计和成功的实验手段而达到的，我们只是刚刚开始探讨细胞通过改变序列库对选择压力产生的反应或细胞容纳额外的序列插入的机理。

第二节　酵母接合型的转变[1~14]

一、酵母的接合型

酿酒酵母(*S. cerevisiae*)有单倍体和二倍体两种形式，在单倍体和二倍体的情况下都能很好的繁殖。各类型之间的转换可通过接合(单倍体孢子融合产生二倍体)和孢子形成(二倍体减数分裂产生单倍体孢子)来实现。进行这些变化的能力由该菌株的接合型所决定。

两种接合型的特征总结于表 11-1。只有当两个单倍体具有不同的基因类型时，接合才有可能发生。并且只有当二倍体是杂合子，因此可产生重组体时，才能产生孢子。

一个单倍体细胞的接合型是由存在于 *MAT* 基因座上的遗传信息所决定的，在这个基因座上，携带 *MATa* 等位基因的细胞称为 a 细胞；而那些携带 *MATα* 等位基因的细胞则被称为 α 细胞。相反类型的细胞能进行接合，相同类型的细胞不能接合。相反接合型的细胞的识别是通过信息素(pheromone)的分泌而实现的。α 细胞分泌一种小分子肽称 α 因子，a 细胞分泌 a 因子，α 因子是一个 13 个氨基酸的肽，a 因子是一个由 12 个氨基酸构成的肽，该肽被法尼基和羧甲基修饰。这些肽皆以前体肽的形式被合成，前体肽经剪接释放出成熟的肽链。

一种接合型的细胞表面携带着相反类型的信息素的表面受体。当一个 a 细胞和一个 α 细胞彼此相遇时，它们的信息素彼此作用，使细胞周期停止于 G_1 期并使细胞产生多种

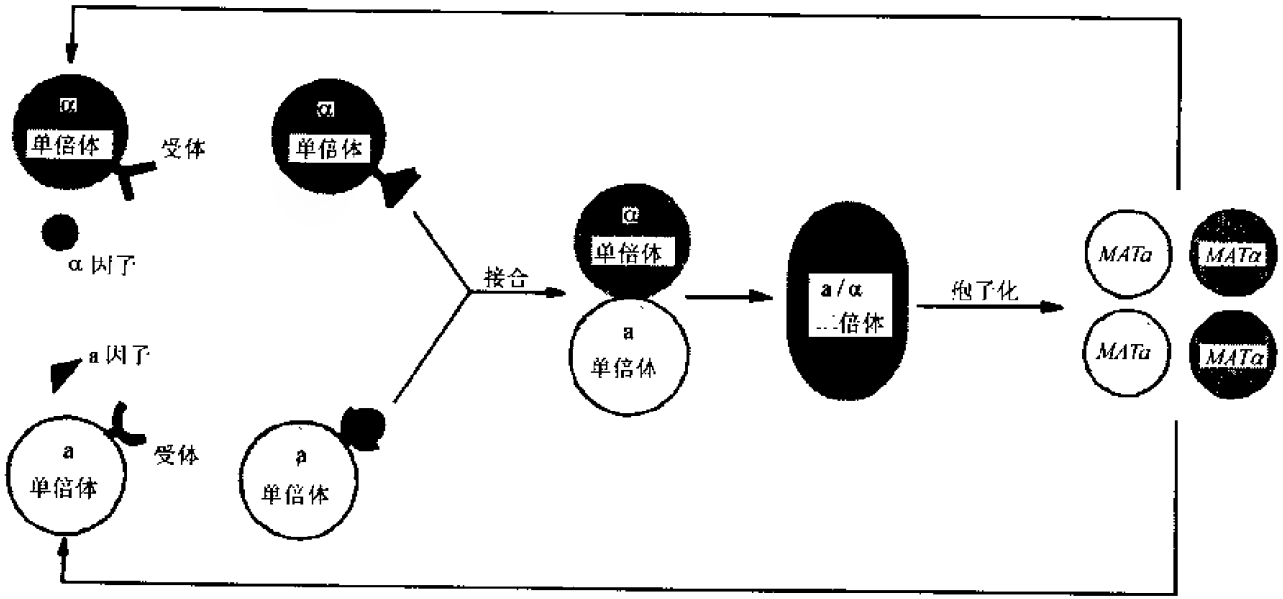


图 11-1 酵母的生活周期

MATa 和 *MATα* 单倍体接合，产生杂合子二倍体。杂合子二倍体孢子化，产生单倍体孢子。

形态上的变化。在一种成功的接合中，细胞周期停止，细胞和核发生融合，产生一个 a/α 二倍体细胞。a/α 细胞携带 *MATa* 和 *MATα* 等位基因，并且与单倍体细胞相比具有完全不同的特征（表 11－1）。特别是 a/α 细胞能生成孢子。图 11－1 所示为单倍体/二倍体正常生活周期。注意：只有杂合子二倍体能产生孢子，纯合子二倍体（a/a 或 α/α）不能产生孢子。

表 11－1　单倍体细胞 *MATa* 和 *MATα* 和二倍体细胞 *MATa*/*MATα* 的比较

项 目	*MATa*	*MATα*	*MATa*/*MATα*
细胞类型	a	α	a/α
接合能力	+	+	−
孢子形成	−	−	+
信息素	a 因子	α 因子	−
受体特征	结合 α 因子	结合 a 因子	−

关于酵母接合途径的信息多是由失去接合能力的 a 和/或 α 细胞突变体的特性而推断出来的。被鉴定的突变基因称为 *STE*（*sterile*，*STE*）。*STE2* 和 *STE3* 基因的突变对一个个体的接合型是特异的。但其他 *STE* 基因的突变使 a 和 α 细胞均失去接合能力，这种情况被解释为对两类细胞来说，在因子与受体相互作用之后所经历的过程是一样的。

接合是一个对称的过程，该过程的开始是一种类型细胞分泌的信息素同另一种类型细胞的表面受体相互反应。在两种接合型的反应途径中所特有的基因是编码受体的基因。任何一种因子和受体的相互作用，都可引发同样的反应途径。在这种共同途径中，那些对某些步骤起删除作用的突变，在两种细胞中产生同样效应。

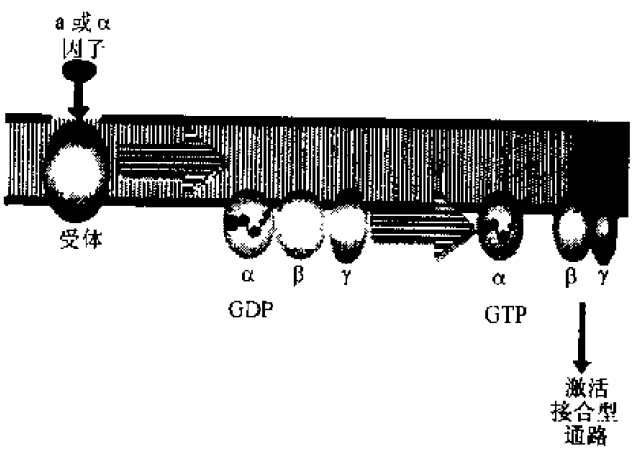


图 11－2　a 或 α 因子与受体结合激活 G 蛋白
a 或 α 因子与相应受体反应，激活 G 蛋白，α 亚单位与 βγ 亚单位分离，βγ 亚单位激活与之相偶联的下一个通路。

接合型反应的起始阶段总结于图 11－2。其组成成分类似于典型的受体-G 蛋白偶联系统。受体是镶嵌的膜蛋白（STE2 是 a 细胞中的 α 受体；STE3 是 α 细胞中的 a 受体）。当任何一个受体被激活时，它与同样的 G 蛋白反应。G 蛋白的三聚体由 α、β 和 γ 亚单位组成。α 亚单位和一个鸟嘌呤核苷酸相结合。在完整的 G 蛋白三聚体中，α 亚单位和 GDP 相连。当信息素受体被激活时，GDP 被 GTP 取代。其结果为 α 亚单位与 βγ 二聚体相分离。G 蛋白亚单位的分离使其能激活与之相偶联的下一个通路的蛋白。

这种通路最常见的机制是利用被激活的 α 亚单位与靶蛋白反应。然而，在接合型的通路上情况有所不同，在该通路中 βγ 二聚体活化下一步骤。现已知 G 蛋白三聚体的各组成成分。它是通过鉴定影响信息素反应的 3 个基因 *SCG1*、*STE4* 和 *STE18* 上的突变实现的。编码 Gα 蛋白的 *SCG1* 的灭活引起信息素反应通路的持续表达（因为 Gα 不能维

持与 Gβγ 以灭活的二聚体的形式存在)。这种突变是致死的,因为它的影响能使细胞周期停止。编码 Gβ 蛋白的 *STE4* 基因或编码 Gγ 的 *STE18* 基因的失活,使接合反应终止,产生不育类型(因为该通路的下个阶段不能被激活)。

该通路随后的各阶段其他的 *STE* 基因已获得鉴定。它们形成一个级联反应(图 11-3)。在该反应中,信号由一个阶段传到下个阶段,最终激活接合所需要的基因。这些基因编码激酶,作用于级联反应的初始阶段;如激酶 STE11 和 STE7 使下一步反应的蛋白磷酸化并因此使之活化。最终转录因子 STE12 被激活;反过来它又激活其产物为接合所需要的基因(类似的级联反应在高等生物中也有发现)。

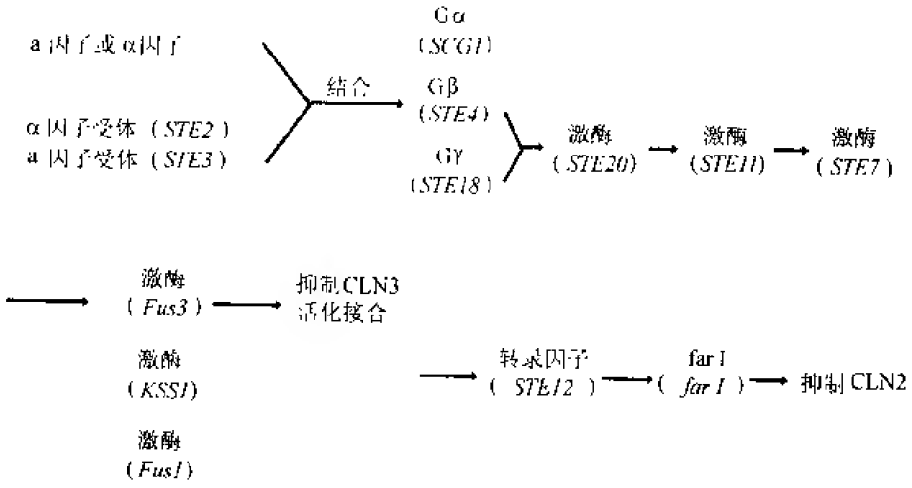


图 11-3 信息素同受体结合产生一种信号沿级联反应依次传递的过程

有几个 *STE* 基因还没有被放到级联反应中,并且所有组成成分的序列尚未确定。有些基因如 *Fus3* 和 *Kss1* 编码激酶,它们可能属于级联反应或存在于它的支路中。然而,原理是清楚的,信息素同受体反应产生一种信号,沿着级联反应依次传递,最终阻抑正常细胞周期所需要的功能并激活接合所需要的功能。

级联反应最终的一些靶物质是这些激酶之一的底物。例如,Fus3 激酶作用于 CLN3,它是细胞周期进行所需要的 3 种 CLN 蛋白的一种。另外一些靶物质则在基因表达水平被控制;例如,另一种 CLN 蛋白,CLN2 是蛋白 farl 作用的靶物质,它的表达是通过转录因子 STE12 激活的。

二、酵母接合型的转变

一些酵母菌株有显著地改变其接合型的能力。这些菌株携带一个显性的等位基因 *HO* 并能经常改变其接合型,其频率可达每一代改变一次。携带隐性等位基因 *ho* 的菌株,具有一个稳定的接合型,其改变的频率大约为 10^{-6}。

HO 的存在会引起一个酵母群体基因型的改变,无论起始的接合型是什么,在很少几代后,该群体即有大量的两种接合型的细胞,从而导致 *MATa* /*MAT*α 二倍体的形成并成为该群体的主体。由一个单倍体群体产生稳定的二倍体群体可被认为是转换存在的理由。

接合型转换的存在表明所有的细胞都含有成为 *MATa* 或 *MATα* 所需要的信息，但只有一种类型获得表达。改变接合型的信息来自哪里呢？类型的转换需要另两个基因座，*HMLα* 为转换形成 *MATα* 型所需；*HMRa* 为转换成 *MATa* 型所需。这些基因座位于携带 *MAT* 的同一个染色体上，*HML* 位于左侧远端而 *HMR* 位于右侧远端。

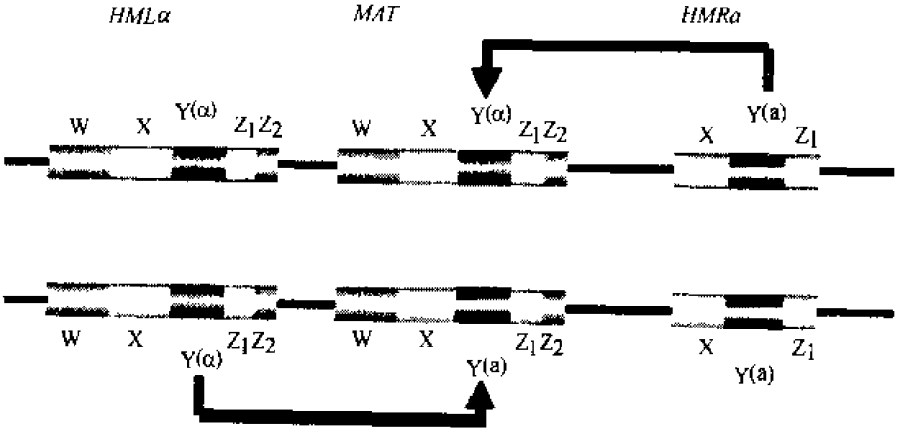


图 11-4 酵母接合型的转换

当沉默匣子取代相反类型的活性匣子时，接合型发生转换。在 *MAT*、*HML* 和 *HMR* 含有同源的 DNA 序列 W、X 和 Z。Y_a 和 Y_α 不同。

图 11-4 为接合型的匣式模型(cassette model)，该模型假定 *MAT* 有一个 α 或 a 型活性匣子(active cassette)，*HML* 和 *HMR* 各含一个沉默匣子(silent cassettes)。通常 *HML* 携带一个 α 匣子，而 *HMR* 携带一个 a 匣子。所有这些匣子都携带编码接合型的信息，但只有在 *MAT* 活化性匣子的信息被表达。当活性匣子被沉默匣子置换时，接合型发生转换，新置入的匣子被表达。

转换并不是相互的，在 *HML* 和 *HMR* 上的拷贝替代 *MAT* 上的等位基因。我们知道这一点是因为通过类型转换 *MAT* 被置换时，*MAT* 上的一个突变被永久的丢失了。它并不与替换它的拷贝进行互换。

如果存在于 *HML* 或 *HMR* 上的沉默拷贝发生突变，转换引入突变的等位基因到 *MAT* 基因座。在 *HML* 或 *HMR* 上的突变拷贝通过无数次转换仍然存留在那里。像复制转座一样，供体元件在受体部位产生一个新的拷贝，而本身仍保留完整无变化。

接合型的转换是具有方向性的过程，在这里只有一个受体(*MAT*)，但有两个可能的供体(*HML* 和 *HMR*)。转换经常涉及到 *MATa* 被 *HMLα* 上的拷贝或 *MATα* 被在 *HMRa* 上的拷贝所置换。有 80%～90% 的转换是 *MAT* 的等位基因被其相反类型的基因所替换，该作用很明显是由细胞表型所决定的。a 表型的细胞倾向选择 *HML* 作为供体，而 α 表型的细胞倾向选择 *HMR* 作供体。

可能获得一些酵母菌株，其沉默匣子的方向被倒转，当它的基因型是 *HMLa* 或 *HMRα* 时，92% 的替换是同源的，在这种情况下，一个 a 匣子被另一个 a 匣子所替换，或一个 α 匣子被另一个 α 匣子替换，这是因为不管沉默基因座的内容是什么，供体匣子的选择是一样的。

确定和转换接合型需要几种基因的参与。除了直接决定接合型的基因外，它们还包括抑制沉默基因表达，转换接合型和执行接合功能有关的基因。

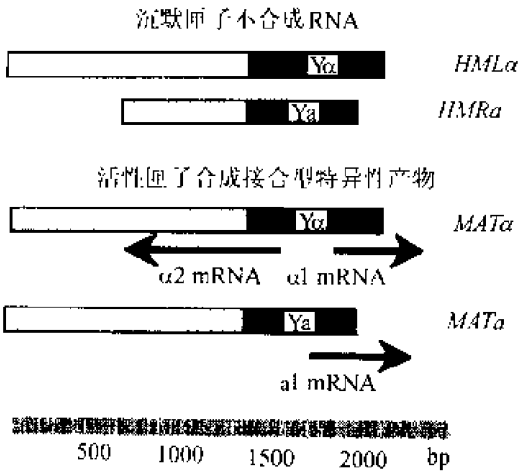


图 11-5 接合型相关基因的结构和转录
沉默匣子与相应的活性匣子有同样序列(*HMRa* 两侧的序列较短)。a 和 α 型的 Y 区序列不同。

通过两种沉默匣子(*HMLα* 和 *HMRa*)与两种活性匣子序列的比较,我们能确定接合型的序列,其基因座的结构总结见图 11-5。每个匣子皆含有位于两侧的共有序列,在 α 和 a 型匣子中心区序列是不同的(称为 Yα 或 Ya 区)。位于该区相应一侧的序列是相同的,虽然 *HMR* 的较短。在 *MAT* 活性匣子由一个位于 Y 区内的启动子启动转录。

MAT 基因座的基本功能是控制信息素和受体基因的表达和其他与接合有关的功能。*MATα* 编码两种蛋白,α1 和 α2。*MATa* 编码一种蛋白,a1。a 和 α 蛋白直接控制不同靶基因的转录;它们的功能通过正或负调控调节。它们的功能在单倍体中相互独立,在二倍体中相互关联。

1) a 细胞 a 特异性基因呈组成型表达,在 α 细胞中表达被抑制。a 特异性基因包括 a 因子结构基因和编码 α 因子受体基因 *STE2*。a 表型与能迅速识别由相反接合型细胞产生的信息素有关。
2) 在 α 细胞中,α 型的特异功能基因被诱导,而在 a 细胞中不表达。这些基因就是 α 因子结构基因和 a 因子受体基因 *STE3*。同样,一种类型的信息素的表达同相反类型信息素的受体表达有关。
3) 单倍体特异的功能基因包括转录信息素和受体的基因,与类型转换有关的 *HO* 基因,和抑制孢子生成的 *RME* 基因。在两种类型单倍体中,它们皆共同被表达,但在 a/α 二倍体中它们的表达被抑制。其结果为在二倍体中 a 特异性和 α 特异性功能不被表达。

现在我们可以由 *MAT* 在单倍体和二倍体细胞中功能表达图解了解到调节因子和它们靶部位的功能。具体见图 11-6。a 和 α 接合型通过不同的机制调节:在 a 单倍体中,a 的接合功能呈组成型表达,在 a 细胞中 *MATa* 产物的功能如果有的话也是未知的。它可能只是用来抑制在二倍体细胞中单倍体的功能。在 α 单倍体中,*α1* 的产物激活 α 接合型所需产物相对应的基因。*α2* 产物则抑制产生 a 接合型有关的基因,它的作用是通过与靶基因上游的操纵基因序列相结合而实现的。在二倍体中,*α1* 和 *α2* 基因产物共同抑制单倍体相关的基因。它们一起识别一个与单独的 *α2* 靶序列不同的操纵基因序列。

α2、a1 和 α1 蛋白调节转录的能力依赖于它们本身的蛋白之间和同其他蛋白之间某些有趣的相互作用。在 a、α 和二倍体细胞中基因调控的类型见图 11-7。

一种被称为 PRTF 的蛋白(不是为接合型所特有的蛋白)参与这些相互作用的过程。PRTF 与一个短的称为 P 框(P box)的共有序列相结合。因为 P 框发现于许多不同位置,表明 PRTF 在基因调节中的作用可能是相当广泛的。在一些基因座上,P 框的作用是活化基因,但在另一些基因座上,PRTF 的作用为抑制某些作用位点上基因的表达。因此,它的作用取决于其他一些与 P 框连邻的位点相结合的蛋白。

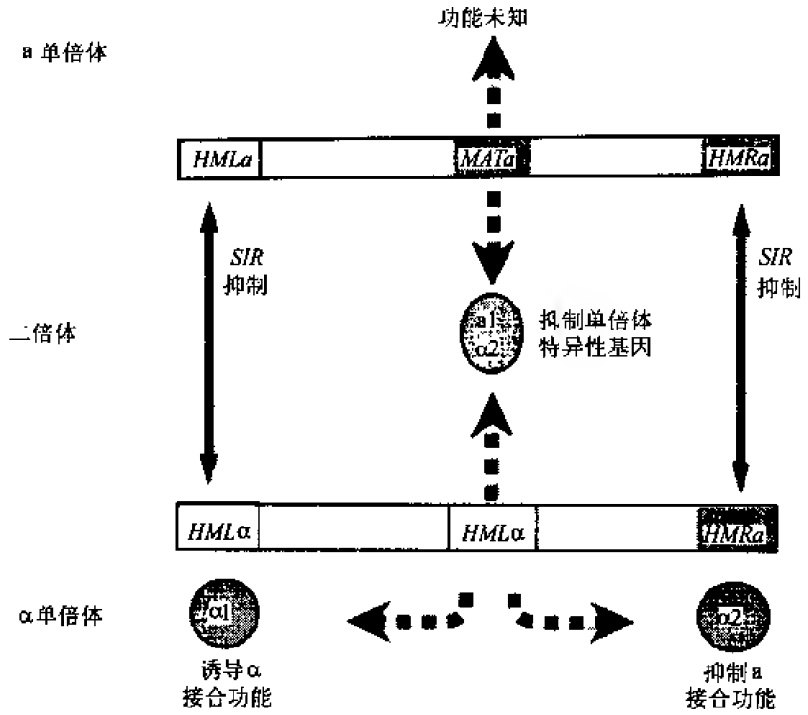


图 11－6　接合功能的调节机制

a 特异性基因可以单独被 PRTF 活化，这能充分保证它们在一个 a 单倍体中表达。

在一个 α 单倍体中，a 基因被 α2 蛋白和 PRTF 的共同作用所抑制。α2 蛋白含有两个结构域，C 末端结构域与 32 bp 的操纵基因共有序列末端一个短的回文成分相结合，然而，这个片段与 DNA 结合并不产生抑制作用。N 末端的结构域是产生抑制作用所必需的，这是因为它负责与 PRTE 发生接触，PRTE 的结合部位是一个位于操纵基因中心的 P 框。实际上，α2、PRTF 以协同的方式与操纵基因相结合。

α 基因的表达需要 α1 活化因子。这是另一种小分子蛋白，由 175 个氨基酸构成。与 α 特异性转录有关的序列位于 UAS 的元件内。共有序列有 26 bp 长并能分成两部分。第一部分是由 16 bp 形成的 P 框，此部位与 PRTF 结合。相连的 10 bp 序列形成 α1 结合部位。α1 因子只有当 PRTF 存在时才能与靶部位结合。两种蛋白都不能单独和它的靶部位相结合，但合在一起则能与 DNA 相结合，这可能是蛋白与蛋白相互作用的结果。

在 a 单倍体细胞，由于 α1 蛋白的缺失，PRTF 不能与 α 特定基因的靶部位结合并使其活化，因此 α 特异的基因不能表达。

α2 蛋白也能和 a1 蛋白共同起作用，它们相结合识别一个不同的操纵基因。此操纵基因与 α2 单独识别的序列有一个大致相同的回文序列，但它较短，这是因为它们之间的序列是不同的。在二倍体细胞中，α2/α1 结合抑制带有这部分特征序列的基因。

三、接合型相关基因的抑制与活化

图 11－5 中的转录过程揭示了一个有趣的特征。*MATa* 和 *MATα* 的转录起始于 Y 区

图 11-7 a 和 α 单倍体和二倍体细胞中基因的调控模式
PRTF、a1、α1 和 α2 分别抑制和活化接合型
相关的基因。

内，但只有 *MAT* 基因座被表达，同样的 Y 区也存在于相应的非转录区（*HML* 或 *HMR*）内。这种情况意味着基因表达的调节不是对于启动子序列相互重叠的某些位点的直接识别所完成。一定有一个位于这些匣子之外的位点能够将 *HML* 和 *HMR* 与 *MAT* 区别。缺失分析表明，*HML* 和 *HMR* 两侧位点是抑制它们表达所需要的。它们就是沉默基因。每个匣子左侧的位点叫 E 沉默基因，右侧的称为 I 沉默基因。这两个部位有两个显著特征：

1）它们的作用像负增强子，例如，它们能够在一定距离（距离启动子达 2.5 kb）和任意方向上发挥功能。

2）它们是和 *ARS*（autonomously replicating sequence，*ARS*）序列相关联的，*ARS* 是复制起点。

我们是否能通过鉴定那些维持这些匣子沉默的基因找到控制这些匣子活化的基础呢？对这些基因的突变研究提供了一个很方便的测定手段。当一个突变使在 *HML* 和

HMR 通常是沉默的匣子被表达时，a 和 α 功能皆被表达，该细胞表现为如同 *MATa*/*MATα* 二倍体细胞。

在几个基因座上的突变终止沉默并导致 *HML* 和 *HMR* 的表达。首先被发现的是 4 个 *SIR*(silent information regulators, *SIR*)基因座。所有 4 个野生型 *SIR* 基因皆是 *HML* 和 *HMR* 处于抑制状态所必需。这些基因座任何一个发生突变，产生一个 *sir*⁻ 等位基因并伴随产生两种效应：*HML* 和 *HMR* 能被转录，并且两个沉默匣子变成转换替代的靶部位。因此，抑制一个沉默匣子的表达并阻止它成为另一匣子替代的受体所利用的是同一个调节过程。

其他抑制沉默基因座的基因，包括 *RAP1*(维持端粒异染色质处于惰性状态所需)和编码组蛋白 H4 的基因。组蛋白 H4 N 末端的缺失或定点突变活化沉默匣子。这些突变的作用能够因在 *SIR3* 引入新的突变或通过 *SIR1* 表达而被抑制，这表明组蛋白 H4 和 SIR 蛋白之间有特异的相互作用。

由这些结果所得出的一个一般结论为 SIR 蛋白影响染色质的结构从而阻止这些基因的表达。由于在 SIR 蛋白的突变对端粒异染色质近端被灭活的基因有同样作用，它似乎表明 SIR 蛋白与组蛋白作用形成异染色质。

当一个匣子处于活化状态时，它有几个高度敏感的部位(图 11-8)。这些高度敏感的部位在沉默匣子上是不存在的，但当引入一个 *sir*⁻ 突变后则所有丢失的敏感部位皆被恢复。在启动子上的高度敏感的部位被认为与基因座表达的能力有直接关系，另一部位和接合型的转换有关，而在沉默子上的那些位点的功能还不知道。

温度敏感的 *sir* 突变体被用于检验 *HML* 和 *HMR* 抑制的建立和细胞复制的关系。携带一个 *sir*⁻ 突变的细胞在高温下温育，*HML* 和 *HMR* 被表达。然后降低温度使 *SIR* 活性恢复。当细胞被阻断于 G_1 相时，*HML* 和 *HMR* 被继续表达，当 DNA 可以进行复制时，抑制反应可得到确立。这表明，同 *E* 元件相关联的 *ars* 序列的利用对确立 *HML* 和 *HMR* 的抑制是必需的。

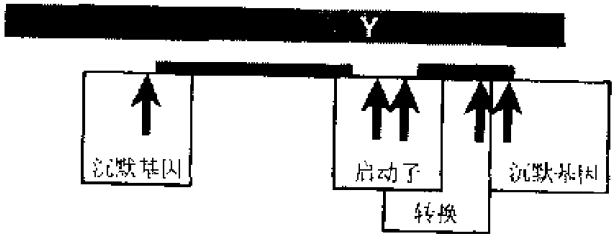


图 11-8 活性匣子上的高度敏感部位
在活性 *MAT* 匣子上，在启动子和两侧连接区有高度敏感部位

一个被称为 ORC 的大的蛋白复合物与 *ARS* 的共有序列相结合，这是复制的引发所需要的。编码这个复合物组分的 *ORC2* 和 *ORC5* 的突变可以终止沉默。这一结果直接证实了在 *ARS* 部位 ORC 的结合是维持沉默所需要的。ORC 的一个可能作用是提供一个起始中心，使沉默效应由此而扩散，这也许是因为它为某种其他成分提供了结合位点。然而，我们还不知道是否 ORC 的单独结合是足够的，或是否复制的某些或所有的后来阶段也是所需要的。

四、接合型转变的机制

接合型的转变是通过受体位点(*MAT*)获得供体(*HML* 或 *HMR*)序列的基因转换所

完成的。通过阻止转换的 *MAT* 上的突变的研究已鉴定了转座所需的位点。由于在 *HML* 和 *HMR* 上缺乏突变位点，表明该过程是单向的。

突变的研究揭示了在 *MAT* 的 Y 区右侧边界处的一个对转换过程十分重要的位点。通过对这些突变点相对于转换部位的位置的分析显示了该边界部位的一些特征(这是通过考查那些尽管突变已经发生但仍然产生了少见的转换的实验结果所得到的)。有些突变位于被置换的区域内(在转换发生后，这些突变消失)，而其余的位于置换区外部(因此会继续阻止转换)。所以位于置换区内部和外部的序列皆是置换过程所需要的。

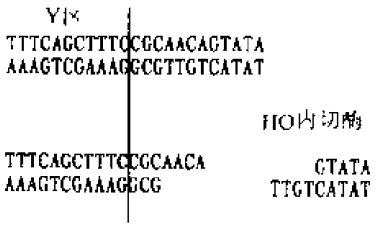


图 11－9　HO 内切酶的作用部位
HO 内切酶在 *MAT* 的 Y 区右侧切割 DNA，产生黏性末端。

转换起始于靠近 Y-Z 边界的双链的断裂，该处刚好是 DNA 酶的敏感部位。该敏感部位因缺乏一个核小体而有可能易于接近。它能被由 *HO* 基因编码的一种内切酶识别。这相当于在体外对 DNA 酶的高敏感性反映出在体内对 HO 内切酶的一种天然的敏感性。

HO 内切酶能在 Y 区紧邻的边界右侧产生双链交错式断裂，酶切后产生一个 4 个碱基的单链末端(图 11－9)。核酸酶对不能转换的突变的 *MAT* 基因座不起作用。缺失分析表明，Y 区接合部周围 24 bp 序列大部分或全部

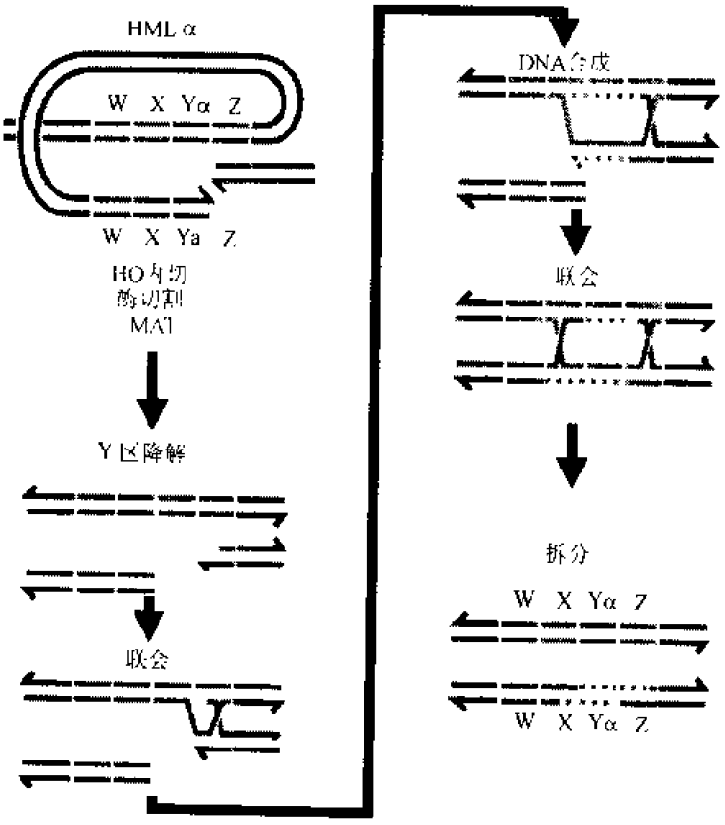


图 11－10　接合型相互转换的模式
原子的替代通过受体(*MAT*)基因座的双链断裂所引发，至涉及到同供体(*HMR* 或 *HML*)基因座 Y 区间的配对，被移出和替代的序列沿伸到 Y 区的左边界。

是在体外切割所必需的，识别位点对于该酶来说是相当大的。有可能此识别序列只见于三种接合型的质子内。

只有 *MAT* 基因座而非 *HML* 或 *HMR* 基因座是内切酶的靶部位。似乎有可能是一种相同的机制维持沉默质子免于被转录，并且也使它们对 HO 内切酶不敏感。这种不敏感性确保转换是单一方向的。

转座反应是由 DNA 的断裂所激发的，图 11－10 所示为从供体和受体区域之间的一般反应的角度对反应的描述。当从 DNA 每个链的相互作用的角度来看，反应过程相当于双链断裂后重组的过程（见重组部分），并且在切割之后的步骤中，需要在一般重组反应中所需的酶的参与。这些基因中的某些发生突变可阻止转换发生。

假设 *MAT* 的游离末端侵入 *HML* 或 *HMR* 基因座并同右侧区配对。*MAT* 的 Y 区被降解直到暴露出左侧的同源区，这时，*MAT* 与 *HML* 或 *HMR* 在左、右两侧皆相互配对。*HML* 或 *HMR* 的 Y 区被拷贝而替代 *MAT* 失去的区域（有可能延伸超过 Y 区本身的界限）。配对的基因座再分离。

像重组的双链断裂模型一样，该过程是由被替代的 *MAT* 基因座开始的。从这种意义上来讲，对供体基因座 *HML* 或 *HMR* 的描述是指它们的最终的作用而不是指这个过程的机制。像复制转座一样，供体部位不受影响但受体的序列发生了变化；和转座不同的是受体基因座发生了替代反应而不是序列的额外增加。

五、*HO* 基因表达的调控

转换是由 *HO* 基因引发的，而 *HO* 基因本身以一种十分有趣的方式被调控，*HO* 基因的转录受几种调控影响：

1）*HO* 受接合型的控制，这是因为它不在 *MATa*/*MATα* 二倍体内合成。从目的论的角度看，当两个 *MAT* 等位基因皆被表达时，进行接合型的转换是无必要的。

2）*HO* 在亲代细胞里而不是在子代细胞被转录。

3）*HO* 的转录也与细胞周期相关，该基因只在亲代细胞 G_1 期末表达。

核酸酶产生的时间可以解释转换和细胞品系的关系。转换只能在分裂的产物中被检测到；两个子代细胞为同样的接合型，是由亲代细胞转换而成的（图 11－11）。这证明了 G_1 期 *HO* 表达的限制确保了在 *MAT* 基因座被复制前进行接合型的转换，其结果为两个子代细胞为新的接合型。

控制 *HO* 转录的顺式作用位点位于 *HO* 基因的上游。与几个调控循环相对应的许多位点中，其中任意一个位点的抑制可阻止 *HO* 基因的转录。

接合型的调控与其他一些单倍体特异性基因相类似。在二倍体中，*α1*/*α2* 阻抑物阻止转录。在上游区有 10 个阻抑物的结合位点。这些位点与它们的共有序列存在不同的差异性；我们不知道其中哪些或多少个为单倍体特异性阻抑作用所需要。

细胞周期的调控是由一个 8 个核苷酸序列的 9 个拷贝所完成的，它们位于－150～－900 之间。与这些序列的一个拷贝相连的基因的表达受细胞周期的调控。除了 G_1 期末期的短暂期间外，与这个序列相连的基因被抑制。它的活性取决于细胞周期中决定细胞分裂的调节因子 CDC28 的作用。

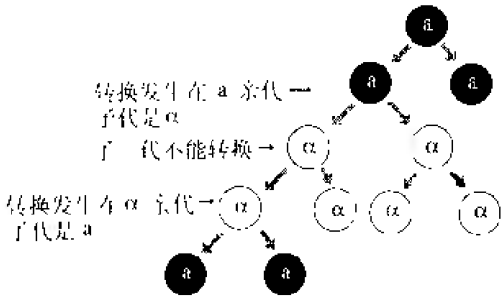


图 11-11 接合型的转换只发生在亲代细胞

接合型的转换只发生在亲代细胞；两个子代细胞有新的接合型，一个子代细胞在能够变为再进行转换的亲代细胞之前必须通过一个完整的循环。

SWI 和 *SIN* 基因产物之间的一些有趣的相互作用与细胞周期和亲代与子代的调控有关。*HO* 的转录需要 *SWI1-5* 基因。它们的作用是阻止 *SIN1-6* 基因产物对 *HO* 基因的抑制。*SWI* 基因首先被发现，是由于其突变体不能进行转换；然后 *SIN* 基因被发现，因为它们能解除特定的 *SWI* 突变体引起的阻断。*SIN* 和 *SWI* 基因之间的反应还不能完全被确定，它涉及到多种途径。

在亲代细胞，发生转录需要 *SWI5*。它的作用是阻断由 *SIN3*、*4* 所产生的抑制。在缺乏这种功能的突变体中，在亲代和子代细胞中 *HO* 同样很好地被转录。该系统作用于上游远端的序列（-1260～-1300 之间）。*SWI5* 本身受细胞周期的控制。它的表达晚于 *HO*。子代细胞先天缺乏 *SWI5* 的功能。所以在它们第一个细胞周期时，如果 *SWI5* 被激活，直到第二个周期它才有机会激活 *HO* 并引起转换。这就解释了子代细胞转换能力的延迟，子代细胞直到已成熟完成一个亲代细胞才能进行转换（图 11-11）。类似的过程，*SWI4* 的产物通过阻抑 *SIN6* 使 *HO* 基因免于抑制。

某些 *SWI* 和 *SIN* 基因不是与接合型特别有关，它们是通用的转录调节元件，许多基因座的表达需要它们的作用。它们包括激活剂复合物 *SWI1*、*2*、*3* 和编码染色体蛋白的基因座 *SIN1*、*2*。它们在接合型表达的作用不是主要的。

第三节　锥虫抗原的变异[1,2,15~22]

一、锥虫的表面抗原

非洲锥虫是一种单细胞的寄生虫。它们交替生存在采采蝇和哺乳动物体内。当采采蝇叮咬哺乳动物时，锥虫可以由采采蝇传递给哺乳动物，也可由患病的哺乳动物传递给采采蝇。

在锥虫的生活周期中，经历了几种形态学和生物化学方面的变化。最显著的生物化学变化发生在可变的表面糖蛋白（variable surface glycoprotein，VSG），它是细胞被膜的主要成分，被膜覆盖着浆膜，它是由 $5\times10^6\sim10\times10^6$ VSG 分子组成的一个单层膜，并且是暴露在细胞表面的唯一的抗原结构。锥虫改变 VSG 的能力对其在采采蝇到哺乳动物的感染循环周期中的生存起主要作用。

如果以一个采采蝇通过叮咬被感染的哺乳动物获得一个锥虫作为循环的开始。锥虫以一种“前循环型”进入到采采蝇的内脏并失去其 VSG。大约 3 周后，它的子代分化成“中期循环型”，并重新获得 VSG 被膜。当采采蝇叮咬哺乳动物时，这种形式被传输到哺乳动物的血液中。锥虫在哺乳动物血流中繁殖。它的子代表达中期循环型 VSG，并持续大约 1 周左右。然后一种新的 VSG 被合成，而且每 1～2 周发生一次转型。

连续产生的每一种 VSG 在免疫学上都是不同的，其结果为呈现给哺乳动物免疫系统的抗原不断地变化，这种转变过程称为抗原变异(antigenic variation)。机体的免疫反应总是落后于表面抗原的改变，因此锥虫能够逃避免疫监视并使之本身能永远存留于寄主体内。VSG 的每一次转变都伴随新一起寄生虫血症，其症状为发烧、皮疹等症状，寄生虫最终侵入到中枢神经系统，在这之后，哺乳动物宿主变得进行性嗜睡最终导致昏迷。

不同的锥虫所能侵染的宿主的范围并不一样。被充分研究的锥虫品种是布氏锥虫(*T. brucei*)的一种变种，它能很好地寄生在实验动物但不能侵染人类。实验室的布氏锥虫能以每次分裂在 10^{-4}～10^{-6}的频率自发的改变 VSG。转换的发生不受宿主免疫系统的影响。实际上，新的变异体随后被宿主选择，因为宿主对旧的 VSG 产生免疫反应，但不能识别和对抗新的 VSG。

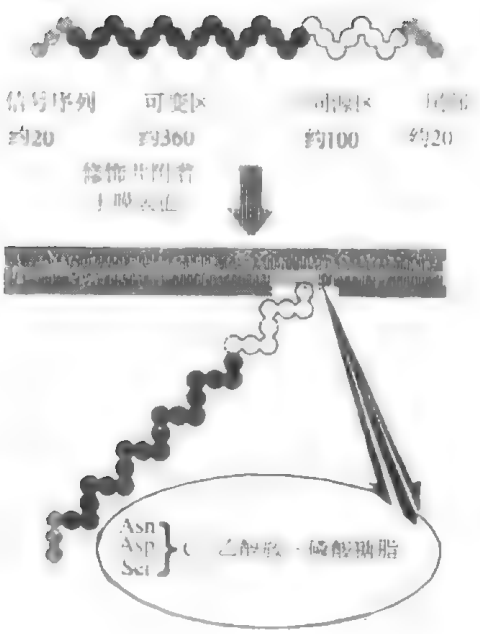


图 11－12　VSG 的结构

VSG 的 N 末端为信号序列和 C 末端疏水的尾部，该末端在成熟后被切除。成熟蛋白的 C 末端通过糖脂和细胞膜共价相连。糖部分位于膜的内部。

VSG 结构的总观见图 11－12。一个新生的 VSG 有 500 个氨基酸长，它有一个 N 末端的信号序列，接着是一个长的可变区，这一可变区提供了唯一的抗原决定簇，而 C 端同源区以一个短的疏水尾部结束。新生的 VSG 的两端经过加工产生一个成熟的 VSG。信号序列在分泌时被切除，疏水尾部在 VSG 达到外表面之前被去除。新的 C 末端与锥虫膜共价结合，依据 C 末端的氨基酸区别同源区的三种类型。

VSG 通过磷酸糖脂与膜结合。因此 VSG 能通过可移去脂肪酸的一种酶的作用而被释放。该反应(已被用于纯化 VSG)在体内可能是很重要的，它能使锥虫表面的一种 VSG 被另一种 VSG 替代。

二、锥虫表面抗原的变异

任意一个锥虫能表达多少种 VSG 呢？目前还不清楚在一个宿主死前是否能达到某种限度。一个锥虫至少能产生 100 种在序列上足够不同的 VSG，使其中任何一种 VSG 的抗体都不能与其他种类 VSG 反应。

在锥虫的基因组中编码不同的 VSG，每个锥虫都携带该株的整个 VSG 基因库。因此，差异性的产生取决于改变基因的表达，由表达存在的一个基因转为表达另外一个基因。

锥虫基因组是由大量的独立单位组成的一个不寻常的结构。除了未知数目的染色体外，它还含有大约 100 个微型染色体，每个含有 50～150 kb 的 DNA。杂交实验鉴别了大约 1000 个 VSG 基因，它们散布在所有不同大小的染色体上。

每个 VSG 由一个基本拷贝基因(basic-copy gene)编码。这些基因可根据它们在染色体上的位置而被分为两类:端粒基因位于距端粒 5～15 kb 内。如果每个端粒有一个的话,则能有 200 多个这样的基因。内部基因位于染色体内(较典型地,它们位于距一个端粒 50 kb 以外的部位)

可以预料作为一个大的基因家族,每个基本拷贝皆显示出相互之间不同程度的相互关联性,这大概反映出它们源于复制和变异。密切相关的基因(引起同样的抗原反应)被称为同源基因(isogenes)。

单一的 *VSG* 基因是如何被选择表达的呢?在某一时刻,只有一个锥虫 *VSG* 基因被转录。活性的基因拷贝被称为表达相连的拷贝(expression linked copy,*ELC*),其位于一个表达位点(expression site),此位点具有如下特征:它位于一个端粒附近。

这些特征表明,选择一个表达基因的途径取决于基本拷贝本身是端粒的还是内部的。能产生 *ELC* 的两种途径见图 11－13。

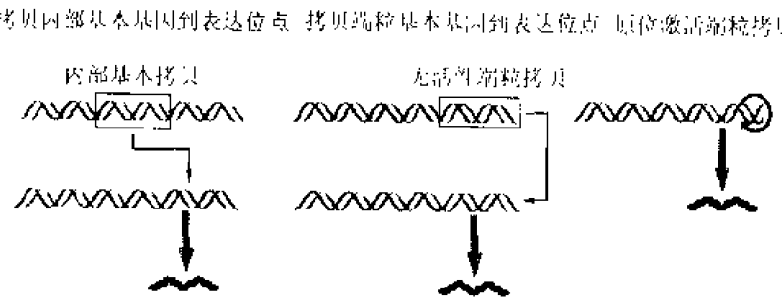


图 11－13 产生 *ELC* 的两种途径
(1)由内部或端粒基本拷贝通过复制转移到表达部位;
(2)通过活化存在于一个可能表达部位的端粒拷贝。

表达位点保持不变,但 *ELC* 被改变。复制转移一个基本拷贝序列替代目前占据表达部位的序列。内部或端粒拷贝皆可直接通过复制到表达部位而被活化。一个基因对另一个基因的替代不干扰活性部位的活性。表达位点被改变,只有当一个序列已经位于一个端粒部位的时候,原位激活才是可能的。当一个端粒部位被原位激活时,先前表达位点的活性必须被终止而新的部位变为表达部位。

内部基本拷贝可能被拷贝到非表达的或表达的端粒部位。这样一个内部基因可通过两个阶段被活化,首先它被转移到一个非表达的端粒,然后该部位被活化。

我们可以通过限制酶图谱了解到涉及活化的基因的踪迹。由 mRNA 得出的表达序列的探针能够检测到与该探针相关的基因。对内部和端粒基本拷贝基因所做的限制酶图谱实验显示不同的结果:

(1)内部基因的活化需要新序列产生　当一个内部基因被活化时,即有一个新片段产生。原来的基本拷贝基因仍然不变;新的片段是通过基因复制到一个新的部位而产生的(通过序列附近的限制酶位点识别因此而产生的一个不同的片段)。新的片段含一个靠近端粒部位的 *ELC* 序列。当基因表达时,此 *ELC* 片段出现,而当基因关闭时,此 *ELC* 片段消失。复制成 *ELC* 是一个内部基本拷贝能够活化的唯一方式

(2)端粒基因的活化能够在原位发生　端粒基因被活化时,基因的数目不发生改

变。限制酶图谱所检测到的基因结构基本不受影响。包含此基因的 DNA 片段的大小可以有轻度的改变，这是由于端粒的长度是经常发生变化的。端粒的基本拷贝也能像内部拷贝一样通过同样复制的方式被活化；在这种情况下，基本拷贝仍然在它的端粒部位，而一个表达相关的拷贝出现在另一个端粒上。

ELC 的形成是通过基因转换产生的，像酵母接合型的转换一样，一个贮备的匣子置换在活性(端粒)部位的匣子。*VSG* 系统是更加多样化的，这是因为存在许多可能的供体部位(同时存在不止一个可能的受体部位)。

几乎所有 *VSG* 类型的转换都涉及到用预先存在的沉默拷贝替代 *ELC*。但也已发现一些例外的情况，在这些情况下，*ELC* 序列不与基因组内沉默拷贝序列中的任何一个匹配。新的序列可以通过一系列的基因转换产生，在这种情况下，不同的沉默基因拷贝的部分短序列被连在一起。这类似于鸡 λ 免疫球蛋白的多样性产生的机制。虽然较少见，这种情况的发生增加了 *VSG* 的多样性。

基因组共有多少表达位点呢？到目前为止，只发现了几个表达位点，这表明只有一部分端粒能发挥这种功能，但有可能任何一个端粒皆可被利用(虽然有一些可能较另一些更倾向于被选择)。

在 *ELC* 的 *VSG* 基因结构是不寻常的(图 11－14)。被转移到 *ELC* 的 DNA 长度是 2500～3500 bp，有大于 1500 bp 的 *VSG* 编码区。大部分附加的长度位于基因的上游。被复制的序列连接到 *ELC* 上的起始位点尚未被精确地测定。

由于成熟的 mRNA 的 5′端带有一个 35 个碱基的序列，该序列是由别处编码的并与新生的 mRNA 的 5′端以反式形式相加，这样使 *VSG* mRNA5′端的分析复杂化了。转录起始和终止(有时也是编码区本身的结束)信号位于与转录区相连的序列。事实上，转录可以在 *VSG* 基因本身上游几千碱基对处起始。启动子已经被测定是在 *VSG* 基因上游 4 kb 和大约 60 kb 处。较远启动子的应用产生一个含有其他基因和活性 *VSG* 的转录物。*VSG* 序列(和其他的基因序列)必须通过对转录物的切割而被释放，在这之后经过剪接的 35 个碱基的前导序列被加到 5′端。

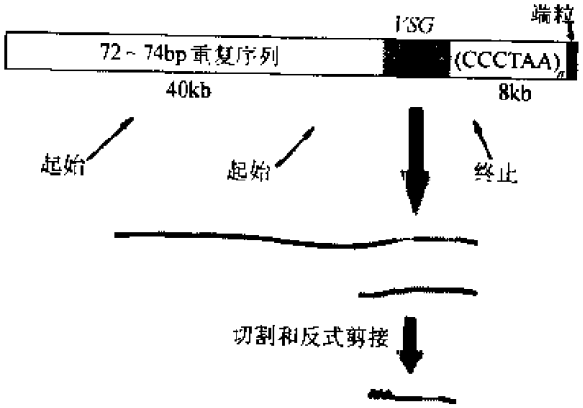


图 11－14 锥虫 *VSG* 基因的 *ELC* 结构

VSG 基因的 *ELC* 拷贝总是位于靠近端粒的部位。基因的上游由大量的 72～76 bp 短的重复序列组成，它们在转换过程中可能提供重组的识别信号。基因的下游为(CCCTA)。重复序列和端粒相连。

转座区的两侧是扩展区，该区不能被限制酶切割。这些“空白”区(barren region)由重复的 DNA 序列组成，它们扩展到 *ELC* 下游 8 kb 左右和上游达 40 kb 处。下游的“空白”区由大量的重复 CCCTAA 序列组成并扩展到端粒部位。上游的序列也可能由 72～76 bp 的重复的序列组成，但它们的性质还不清楚。“空白”区的存在已经成为通过克隆鉴定 *ELC* 基因的障碍。

一个表达部位的活化涉及到转录单位上游的某种变化。当一个 *ELC* 的启动子被另

一个启动子替代时，*ELC* 被灭活和被活化的能力并不改变。因此，启动子不受反式作用因子调控。调控过程可能是起始于上游区的染色质结构上的改变。

在感染的过程中，*VSG* 基因的表达的顺序是不规则的，但并不是完全随机的。这可能是保证锥虫存活的一个重要特征。如果 *VSG* 基因以一个预先确定的程序被表达，则一个宿主可以通过产生抗早期成分的一种反应而消除感染。*VSG* 产生的不可预料性的需要，也许是促成该系统进化成带有很多供体序列和多个受体的原因。

抗原变异不是锥虫独有的现象。赫氏螺疏螺旋体（*Borrelia hermsii*）可引起人的回归热和其他动物类似的疾病。疾病的名称反映出它的不稳定过程：整个发病期由许多缓解期所隔开。当发烧时，血液中发现有螺旋体；在缓解期它们消失，这是因为宿主产生了对抗螺旋体的特异性抗体。

像锥虫一样，螺旋体通过改变一种表面蛋白而存活，该蛋白称为可变的主要蛋白（variable major protein，VMP），*VMP* 的变化和基因组的重排有关。活性 *VMP* 位于一个线性质粒的端粒附近。我们还不知道存在多少编码的变异体和用于改变它们表达的机制。然而有趣的是真核锥虫和原核螺旋体都依赖于抗原变异作为逃避免疫监视的方式。

第四节　DNA 的扩增[1,23~29]

一、基因组某些序列的扩增

真核基因组有能力容纳外源或内源的附加序列。外源序列通过转染（transfection）引入到细胞内，内源序列可以通过一已存在序列的扩增（amplification）产生。这两种情况共有的特征是附加的序列很可能采取串联排列的形式，此结构含有一重复单位的许多拷贝。位于重复单位中的基因拷贝并不是每个皆被表达，但是表达倾向于随着拷贝数目的增加而增加。

通过转染或扩增产生的多拷贝的串联排列在细胞中以两种形式存在。如果采取染色体外单位的形式存在，它将以一种不规则的方式遗传；动物细胞没有与细菌在细胞分裂时对质粒进行均匀分离的能力相当的功能，因此整个单位被高频率丢失。然而如果序列被整合到基因组中，它变为遗传的一个成分并像任何其他的基因组序列一样被遗传。

内源性序列的扩增是通过选择抗某些试剂的细胞而引起的，扩增最好的例子起因于对某些培养细胞株中氨甲蝶呤（methotrexate，MTX）的加入，该试剂阻断了叶酸的代谢，对它的抗性可通过改变二氢叶酸还原酶（dihydrofolate reductase，DHFR）活性的突变而获得。改变酶自身的一种替代方式为增加酶的总量，这种酶量增加的原因是因为 *dhfr* 结构基因的数目扩增了。扩增的频率大于自发性点突变率，通常的范围为 10^{-4}～10^{-6}。类似的过程现在已在 20 多种基因中观察到。

这些系统的大部分的共同特征是高度抗性细胞不是通过单一阶段而获得，相反，是通过细胞对逐渐增加毒剂适应而产生的。因此，基因的扩增可能需要几个阶段。扩增一般只出现在两个 *dhfr* 等位基因中的一个，并且对氨甲蝶呤抗性的增加是通过该基因座扩增程度的进一步提高而完成的。

在一个抗氨甲蝶呤的细胞株内 *dhfr* 基因的数目大约在 40～400 之间，这取决于选择

的强度和细胞株。mtx^r 细胞株可分成两类，其差别见于当生长在缺乏氨甲蝶呤的培养基中，对高水平 *DHFR* 活性的选择压力消除时它们的反应不相同。在稳定的细胞株，扩增的基因被保留，因为它们位于染色体上，通常位于被单一 *dhfr* 基因占据的部位。而另一染色体保留它原来的单拷贝 *dhfr*。在不稳定细胞株，当选择压力被解除时，扩增的基因至少部分丢失，因为扩增的基因存在于染色体外。

基因扩增会在染色体上产生可见的影响。在稳定株上，*dhfr* 基因座以均匀染色区的形式显现(homogeneously staining region，HSR)。当对染色体进行处理后(如 G 带处理)，则会发现不止一个缺乏任何染色带的区域，这即是 HSR 名称的由来。这种变化表明，在染色体一些区域的带间部分发生了扩增。

在不稳定的细胞株内，携带 *dhfr* 的染色体没有发生变化。然而，会发现大量的被称为双微染色体(double-minute chromosomes)的成分的出现(见图 11－15)。在一个典型的细胞株内，每个双微染色体携带 2～4 个 *dhfr* 基因。双微染色体似乎是自我复制的，但它们缺乏着丝粒，所以不会被吸附到有丝分裂的纺锤体上，因而有规律地分离，其常常从子细胞内丢失。尽管它们有命名，但双微染色体实际上被看作是染色体外成分。

双微染色体的无规律遗传解释了这些株对氨甲蝶呤的抗性的不稳定性。在细胞分裂时，双微染色体持续的丢失，在氨甲蝶呤存在时，含较少的 *dhfr* 基因数目的细胞将会死亡，只有那些保留有足量双微染色体的细胞将存活下来。

双微染色体的存在减少了细胞增殖的速率。因此，当选择压力被除去时，缺乏扩增基因的细胞则具有优越性，它们以更快的速率产生后代，并很快取代原来的存活群。这就解释了为什么只有细胞生长在氨甲蝶呤的培养基中，扩增状态的细胞株才被保留。

由于双微染色体的不规则分离，它们在细胞中的拷贝数可以相对较快地增加，这是因为在每次分裂中某些细胞可获得多于它们平均应当获得的 *dhfr* 基因。随着氨甲蝶呤水平的升高，细胞所携带的拷贝数亦增加。双微染色体的性质解释了 mtx^r 条件下的逐步进化和在不稳定细胞株内 *dhfr* 基因水平不断的波动性。

氨甲蝶呤抗性的长期选择使稳定和不稳定的细胞株得以产生。基因扩增的起始阶段是什么？在短期的选择后，大多数或所有具有抗性的细胞是不稳定的。染色体外拷贝的形成很明显比染色体内的扩增较常见。在这个过程的最早期，在双微染色体或任何染色体的变化可检测到之前，即可发现扩增的 *dhfr* 基因以很小的染色体外单位的形式存在。这表明，抗性的获得常常是由于染色体外重复单位的产生。

扩增区大大长于 *dhfr* 基因本身。基因大约有 31 kb 长，而在染色体 HSR 上的重复单位的平均长度是 500～1000 kb。每一个细胞株内的扩增区的长度是不同的，一个双微染色体含有的 DNA 量似乎是在 100～1000 kb 的范围内。

染色体外的拷贝是怎样产生的呢？我们知道，它们的产生没有使原来染色体的拷贝丧失。这有两种可能性：在 *dhfr* 基因附近发生附加的复制循环，接着在这些附加的拷贝之间发生非同源性重组。或者该过程起始于等位基因之间的非同源性重组，额外的拷贝可通过某种类似重组的过程由染色体中释放出来。根据该过程的性质，能够产生含有一个或几个拷贝的染色体外的 DNA 分子。如果双微染色体含有环状 DNA，在任何情况下，它们之间的重组都很可能产生多聚分子(图 11－15)。

通过对一特殊的不稳定的细胞株的研究，揭示了双微染色体生存有关的一些信息，在

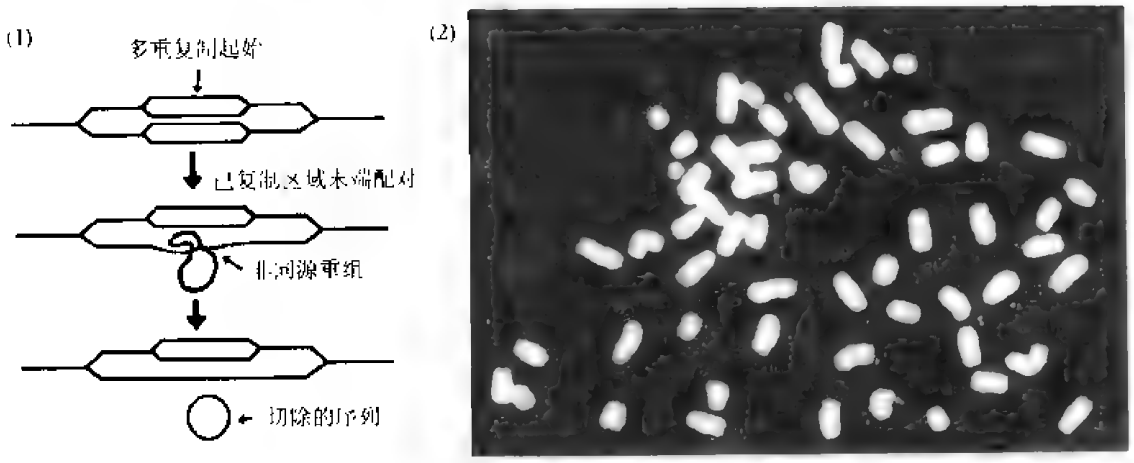


图 11　15　双微染色体及其形成机制

(1)含 *dhfr* 基因的双微染色体形成机制；(2)以小白点形式存在的双微染色体。

此细胞株中，被扩增的基因编码一种突变的 DHFR 酶。突变的酶不存在于原来二倍体的细胞株中(因此突变一定产生于扩增过程中)。尽管被扩增的基因在数目上不同，但这些细胞只表达突变的酶。这说明野生型染色体的基因不能连续地产生大量的新双微染色体，因为这些扩增的拷贝会产生正常酶。

一旦扩增的染色体外基因已经产生，细胞状态的变化即通过这些基因而不是通过原来染色体的拷贝介导。当氨甲蝶呤被移除时，细胞以通常的方式失去双微染色体。再暴露在该试剂后，正常的基因被扩增产生一个新的双微染色体群体。这表明没有突变基因的染色体外拷贝整合到染色体中。

这些结果的另一个显著意义是突变株的双微染色体只携带突变基因，因此如果每个双微染色体含多个 *dhfr* 基因，它们一定皆是突变型。这显示多拷贝的双微染色体可由一个染色体外的基因产生。

一个主要的问题是染色体拷贝数的增加是通过染色体外拷贝的插入(整合)还是通过一个独立的机制产生的。我们不知道是否染色体内的扩增仅仅是以较低的频率从头开始的或需要染色体外扩增的出现作为一个中间步骤。被扩增的基因所采取的形式受细胞的基因型影响，一些细胞株倾向于产生双微染色体，而另一些细胞株较容易产生 HSR。

扩增的类型也取决于所涉及的特殊的基因座。扩增的另一种情况由转氨甲酰酶(transcarbamylase)抑制剂的抗性的产生所提供，这种情况下发生 *CAD* 基因的扩增(CAD 蛋白具有 UMP 合成途径中头三种酶的活性)。被扩增的 *CAD* DNA 总是存在于染色体内。在这种情况下，被扩增的基因存在于几个分散的扩增区，并常常涉及到一个以上的染色体。

二、引入外源性 DNA 到活细胞和动物体内

引入外源性供体 DNA 到受体细胞的过程称为转染。转染实验始于加入制备的中期染色体到细胞的悬浮液中。细胞以相当低的效率摄入染色体并产生不稳定的变异体。在

该程序中，完整的染色体很少能够得以保存；受体细胞常常获得供体染色体的一个片段（由于它缺乏着丝粒，所以是不稳定的）。在极少的情况下，供体物质可整合到受体的染色体上，从而产生稳定的细胞株。

当纯化的 DNA 被加入到受体细胞的制备物中时，可以获得类似的结果。然而，用纯化的 DNA 代替随机的染色体片段可以使一个特殊的序列被加入到受体细胞。用纯化的 DNA 转染可产生稳定的和不稳定的细胞株，前者相对占优势（这些实验可直接类比于细菌的转化过程，但被定名为转染是由于在历史上转化被用来描述那些使真核细胞无限制生长的变化）。

不稳定的转染，有时称为瞬时转染（transient transfection），表明转染的 DNA 以染色体外的形式存在，在稳定株中转染的 DNA 插入到基因组中，被转染的 DNA 在两种情况下皆可被表达。然而，由于转染的频率低，有必要利用可被选择的供体标记来鉴别被转染的受体细胞。很多转染实验利用容易被检测的酶的功能作为标志，但原则上任何能被检测的标志皆可被利用，这就允许与某些表观形态相对应的基因得以分离。最典型的转染的细胞可通过获得转化的表型（形成肿瘤）而被选择。利用此方法已分离到几种细胞癌基因。

已经证明，带多种选择标志的共转染可提供有关转染过程的信息，并扩展了我们应用这种技术探讨问题的范围。用于这种实验的一个常见的标志是 *tk* 基因，它编码胸腺嘧啶激酶，该酶催化胸腺嘧啶核苷三磷酸合成过程的一个重要步骤。

当用一个含有 tk^+ 基因的 DNA 片段和 ΦX174 基因组的 DNA 混合物转染 tk^- 细胞时，所有的 tk^+ 转化子皆含有两种供体序列。这是一个有用的实验结果，因为它通过与一个选择性标志的共转染使一个非选择性标志程序化地引入到细胞中。

tk 和 ΦX174 序列的排列在每一个被转染的细胞株中是不相同的，但它们在细胞株的繁殖过程中保持不变。供体序列经常以许多拷贝的形式存在，其数量随细胞株的不同而异。回复子同时失去 ΦX174 和 *tk* 基因序列。在选择性压力的作用下，被转染的序列的扩增会引起所有供体序列的拷贝数同时增加。因此这两种类型的供体序列在转染过程中发生相互联接并在此后经历同样的命运。

为了进行转染实验，加到受体细胞中的 DNA 量可通过加入一种额外的“载体 DNA”（一种 DNA 制品，常常来自鲑鱼的精子）而增加。经证明，被转染的细胞发现有位于选择序列两侧的载体 DNA 序列。因此，转染似乎是被一个大的结构单位所介导，其组成为一个存在于供体制品中所有序列相互连接的长序列。

被选择性标志所转染的细胞的回复子失去了所有的这些序列，因此似乎被转染的细胞只获得一个单一的大的结构单位。这一结构单位由供体序列相连接而形成，转染结构单位的长度大约为 1000 kb。

由于供体单位的这种长度，我们不能由印迹实验来确定它是否在物理上连接到受体的染色体 DNA 上（相关的末端片段的相对比例太小）。看起来可能是第一阶段是建立不稳定的染色体外的单位，接着通过整合到染色体中而获得稳定性。

原位杂交实验可被用来显示转染细胞中有供体整合到细胞的染色体中。任何一个细胞株只有单一的整合位点，并且每个细胞株的整合位点皆不相同。这表明对整合部位的选择可能是一个随机过程，有时它和染色体重排相关。

外源性DNA常常与整合位点的DNA没有任何序列的相关性。整合涉及到加入的DNA和基因组的随机位点之间的一种非同源性重组的过程。重组过程可能由在染色体上引入一个双链的断裂而引发，该引发过程可能是通过DNA修复酶的作用而实现的，此酶可被外源性DNA的游离末端所诱导。

转染技术一个令人振奋的发展是它们可用于将基因引入到动物体内，由于外源性DNA的加入而获得新的遗传信息的动物称为转基因动物。直接注入DNA的方法可被用于鼠的卵子，携带所需基因的质粒可被注入到卵母细胞核或受精卵的原核，随后受精卵被植入到假孕鼠的子宫内。受体鼠出生后，可检测是否已经获得了外源性DNA，如果有的话，它是否被表达。

对任何转基因动物，我们要问的第一个问题是它含有多少外源性DNA拷贝，这些拷贝位于什么部位，它们是否存在于生殖细胞并以孟德尔的方式遗传。这些实验的通常结果是：被注射的鼠小部分（例如，约15%）携带转染序列。通常，质粒的许多拷贝以串联的形式被整合到染色体的单一位点，拷贝的数量为1～150。并且这些拷贝可能以孟德尔基因座（Mendelian locus）的形式遗传给后代。

重要的一点是通过转基因动物的实验可以探讨基因的独立性以及和它们所在部位相关的区域效应。如果我们选择一个基因，并包括它两侧的含有已知调控元件的序列，在基因组内它能不受所在位置的影响而独立地表达吗？换句话说，调控元件是独立地行使其功能还是基因表达还受其他因素的控制，例如，对位于染色体结构域的一个适当的位置需求。

被转染的基因是以适当的发育特性被表达的吗？一般的原则似乎是存在一个适当调控的合理模式：转染的基因一般在适当的细胞里并且在通常的时间被表达。然而，有一些例外，在有些情况下转染的基因在一个不适当的组织内被表达。

在被注射的鼠的后代，供体基因的表达是相当不同的；它可以完全消失，有一些减少，或甚至增加。即使在原来的亲代，基因表达的水平同串联的插入基因的数目也不相关。也许只有一些基因具有活性。除了多少基因拷贝能够被活化的问题外，影响调控的一个参数可能是基因数目和调节蛋白之间的关系：大量的启动子数有可能冲消有限数量的调节蛋白。

基因表达的变化和什么有关呢？对于转染的基因（也适用于被整合的反转录病毒基因组），一种常常被讨论的可能性是整合的位点具有重要影响。如果一个基因被整合到一个活性区内，它可能被表达，但如果被整合到染色质的其他区域内，它可能不被表达。另一个可能性是发生外遗传的修饰；例如，甲基化类型的变化可能与活性变化有关。另外，在亲代具有活性的那些基因，在子代有可能被缺失或得到扩增。

注入基因的作用的一个显著的例子来自一种特殊的转基因鼠，该鼠来源于用*MT*启动子和大鼠生长激素的结构基因的融合序列所转染的卵细胞。某些转基因鼠生长激素的水平是正常鼠的几百倍，此种鼠可长到大约正常鼠两倍的大小。

致癌基因的引入可导致肿瘤的形成。含有SV40早期编码区和调控元件的转基因鼠只在某些组织表达大T和小t抗原的病毒基因（通常是在脑、胸腺和肾脏；T/t抗原是病毒早期区域编码的蛋白质的不同剪接产物；它们具有转化培养细胞为肿瘤生成表型的能力）。此种转基因鼠的存活期常常小于6个月，这是脑肿瘤的形成所引起的，有时肿瘤也

可发现于胸腺和肾脏。不同的癌基因可被用于产生具有不同类型癌症的鼠，这使制造各种模型系统成为可能。例如，在一个活性启动子的控制下的 *myc* 基因的引入可引起腺癌和其他肿瘤的形成。

在性细胞中，缺陷的基因能否利用转基因技术用有功能性的基因代替？一个成功的例子是对性功能低下(hypogonadal，hpg)鼠的治疗。*hpg* 鼠有一个大于 30 kb DNA 片段的缺失，这使编码促性腺激素释放素(gonadotropin-releasing hormone，GnRH)和与 GnRH 相关的多肽(GAP)的多蛋白(polyprotein)前体基因的末端部分丢失，其结果使鼠失去生育能力。

当一个完整的 *hpg* 基因通过转基因技术引入到鼠体内之后，它即在适当的组织中得到表达。通过实验引入基因到 *hpg/hpg* 纯合子突变鼠中，由此产生的鼠是正常的。这表明，在正常的调控机制下，转基因与正常的等位基因的表达是没有区别的。

目前，应用这种技术治疗遗传缺陷病的障碍是基因必须引入到上一代的生殖细胞之中，它的表达能力是不可预料的，并且所转基因有可能只在一小部分转基因动物中被充分表达。大量的基因能被引入到生殖细胞中，它们的表达不稳定，而基因的过表达有可能对动物有害。

例如，在 hpg 鼠的实验中，被注射的 250 个鼠卵只产生 2 个带有完整的 *hpg* 基因的转基因鼠，每个转基因鼠含有 20 多个该基因的拷贝。在 48 个转基因鼠的后代中，只有 20 个保留有转入基因的特性。转入的基因能替代缺失的内源性 *hpg* 基因。因此，通过转基因途径产生的基因替换只在限定的条件下才是有效的。

DNA 直接注入的缺点是多拷贝的引入及其不稳定的表达，并且因为有可能使宿主 DNA 中产生序列的重排，常常很难使克隆插入正确位点。另一种方法是用一个携带有供体基因的反转录病毒载体，一个原病毒拷贝插入到一个染色体的位点，而且不引起任何宿主 DNA 的重排。也可以在细胞的不同发展阶段对其进行处理，这样能针对一种特殊的体细胞组织，但是感染生殖细胞是很困难的。

制造转基因鼠的一个有效的技术是采用胚胎干细胞(embryonic stem，ES)，它来自于鼠的胚泡(产生于卵细胞植入子宫之后，胚胎发育的早期阶段)。DNA 以通常的方式转染 ES 细胞(常常通过微注射或电穿孔)。通过利用一个携带附加序列(例如，一个药物抗性标志或某种特殊的酶)的供体，使已经获得供体的 ES 细胞的选择成为可能。另一种方法是可利用 PCR 技术确定供体 DNA 是否成功地插入被转染的 ES 细胞。通过这些方法，可获得一个 ES 细胞群体，其中携带标志的细胞占很高比例。

这些 ES 细胞然后被注射到受体胚泡中，它们能参与胚泡正常发育的能力构成了该技术的基础。胚泡被置入到代孕母鼠，经过一个适当的时间即发育成一个嵌合鼠。嵌合鼠的一些组织将衍生于受体的胚泡细胞，其他的组织将衍生于被注入的 ES 细胞。衍生于受体胚泡和衍生于被注入的 ES 细胞的成年鼠组织的比例是相当不同的。如果利用一个可见的标志(如表皮颜色基因)，则每种类型细胞的组织区域可用肉眼观察到。

为了鉴定 ES 细胞是否能分化成生殖细胞，可用嵌合鼠与一个缺乏供体特性的鼠进行交配。任何有供体特性的后代一定是从注射的 ES 细胞分化的性细胞衍生而来。通过这种方法，一个完整的鼠可由一个初始的 ES 细胞产生。

这些技术的进一步发展使获得一个同源性重组体成为可能。同源性重组的一个特殊

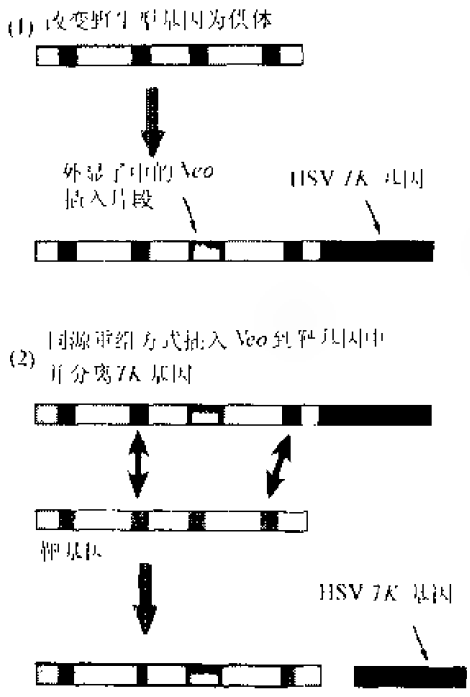


图 11-16 同源性重组体的构造
通过在一个野生型的基因的外显子内插入 *Neo* 基因，在一端加入 *TK* 基因，通过同源性重组插入该基因到靶部位并且分离 *TK* 基因。

用途是破坏内源性基因，如图 11-16 所示，一个野生型基因通过插入的一个带有标志的外显子序列而被修饰；最常用的标志序列是 *Neo* 基因，它可提供一种对 G418 药物的抗性。此外，另一种标志被加入到该基因的一端，例如，枯草病毒的 *TK* 基因。当此 DNA 被引入到一个 ES 细胞时，它可以通过同源或非同源性重组而整合到基因组中。非同源性重组插入整个序列单位，包括末端的 *TK* 序列。但是，同源性重组需要同源序列交换，其结果为末端的 *TK* 序列被丢失。发生同源性重组的细胞可通过获得 *Neo* 抗性和缺乏 *TK* 活性而被鉴别（*TK* 基因可引起对药物 gancyclovir 敏感）。如果用一种选择标志如 *TK* 不方便的话，可通过 PCR 测定而简单地筛选不存在 DNA 末端部分的细胞。同源性重组的频率大约为 10^{-7}，这可能小于全部重组事件的 1%。

外显子中 *Neo* 基因的存在干扰转录，因此产生一个失效的等位基因。特殊的靶基因可以通过这种方式敲除（knock out）。一旦获得带有一失效等位基因的鼠，即可通过培养后产生纯合子鼠。这是十分有用的技术，通过它可以研究一个特定的基因是否十分重要的，和它的丧失使动物的哪些功能发生紊乱。

引入新 DNA 序列的一个先进的方法是通过利用果蝇 *P* 元件来实现的。一个携带所需基因的缺陷型 *P* 元件同一个完整的 *P* 元件一起被注射到胚胎的前胚盘中。完整的 *P* 元件提供转座酶，它不仅识别完整 *P* 元件本身的末端而且也识别缺陷 *P* 元件的末端，这样，任一种或两种元件可被插入到基因组中。

只有在 *P* DNA 末端之间的序列能被插入，在其两侧的序列不是转座子的组成部分。这种技术的优点是在此过程中，只有一个单一的元件被插入，因此，转基因的果蝇通常只携带一个外源基因的拷贝，这对分析它的属性有很大的帮助。

应用这种方式引入的几种基因皆显示同样的属性。它们只在适当的组织和发育的适当阶段被表达，并且表达与插入的部位无关。

所以对果蝇来说，基因座本身可能含有调节基因表达所需的全部信息并且可能相对的不受外界干扰。

这些实验表明，由培养细胞到动物皆可作为研究对象来考察调控的特性。引入 DNA 到基因型（genotype）的可能性允许我们对其加以改变：加入在体外经过特别修饰的新基因，或灭活存在的基因。这样使描述组织特异基因表达的特性成为可能。最终，我们可以利用定位的方式，替换基因型中的缺陷基因。

（张玉静）

参 考 文 献

[1] Lewin B. Genes. Oxford University Press, 1997. 989
[2] David R F. L. Genetic Recombination. Blackwell Science Ltd, 1996
[3] Haber J E. Trends Genet, 1992. 8:446～452
[4] Klar A J S. Mobile DNA. Washington: American Society for Microbiology, 1989. 671
[5] Nasmyth K. Ann Rev Genet, 1982 16:439～500
[6] Nasmyth K et al. Science, 1987. 237:1162～1170
[7] Kurjan J. Ann Rev Biochem, 1992. 61:1097～1129
[8] Laurenson P et al. Microbiol Rev, 1992. 56:534～560
[9] Kurjan J. Ann Rev Genet, 1993. 27:147～179
[10] Beltort M. Ann Rev Genet, 1990. 24:363～385
[11] Strathern J N. Genetic Recombination. Washington: American Society for Microbiology, 1988. 445～464
[12] Hicks J et al. DNA Insertion Elements. New York: Cold Spring Harbor Laboratory, 1977. 457～462
[13] Strathern J N et al. Cell, 1982. 31:183 - 192
[14] Bender A et al. Cell, 1986. 47:929～937
[15] Donelson J E. Mobile DNA. Washington: American Society for Microbiology, 1989. 763～782
[16] Swanson J. Mobile DNA. Washington: American Society for Microbiology, 1989. 743～762
[17] Boothroyd J C. Ann Rev Microbiol, 1985. 39:475～502
[18] Donelson J E. Microbiol Rev, 1985. 49:107～125
[19] Borst P. Ann Rev Biochem, 1986. 55:701～732
[20] Pellicer A et al. Science, 1980. 209:1414 - 1422
[21] Schimke R T et al. Cold Spring Harbor Symp. Quant. Biol, 1981. 45:785～797
[22] Stark G R et al. Ann Rev Biochem, 1984. 53:447～491
[23] Toledo F et al. Current Biology, 1993. 3:255～462

第十二章　免疫多样性产生的机制

第一节　概　　述[1-4]

免疫系统对抗原识别的能力依赖于B细胞产生的抗体和T细胞产生的抗原受体(T-cell receptor,TCR)。虽然T细胞和B细胞识别抗原的方式相当不同,但两种细胞群体皆能识别相当广范围的抗原。尽管抗体和T细胞受体之间有很大差别,但每种类型分子产生多样性的细胞和分子过程是十分相似的。本章集中阐述免疫系统产生大量的,具有不同的抗原特异性的抗体和T细胞抗原受体的方式。

免疫球蛋白具有显著的多样性,能够提供足够的不同结合位点,以识别环境中数以千万计的抗原分子类型。每类免疫球蛋白都有一个特征性的效应区,如IgE能与肥大细胞表面的FC受体结合,而IgG能与吞噬细胞结合。据估计,一个人产生的不同类型抗体的数目比体内其他种类蛋白质的总和还多。事实上,我们体内产生的各种抗体数多于基因组基因总数。免疫球蛋白的多样性是如何产生的呢?在过去的若干年中,关于抗体如何产生的理论,已经发生了很大改变,但令人惊奇的是,Ehrlich在1897年提出的侧链假说(side-chain hypothesis),十分接近当前的理论(图12-1)。

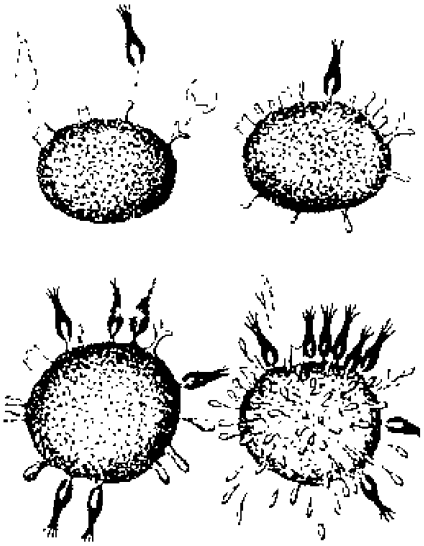


图12-1　Ehrlich的侧链理论

Ehrlich提出,抗原和前B细胞(preformed B-cell)表面的受体结合,激发该细胞产生和分泌更多的这种受体。如图中指出,他认为一个细胞能产生结合多于一种类型抗原的抗体,但很显然,是他首先提出了克隆选择的理论,以及在接触抗原之前,免疫系统能产生相应受体的想法。

除了他认为同一细胞上可有几种不同的受体侧链之外,他的抗原诱导选择的观点,很接近目前克隆选择的理论。

一、免疫球蛋白形成的理论

Ehrlich的观点提出后,情况变得复杂化了。当时合成了许多新的有机化合物,而且Landsteiner证明免疫系统能针对每种新化合物产生特异性抗体。那时认为,通过自然选择,免疫系统能保持针对所有这些新的、人工合成的化合物的抗体全部基因是不可能的。这样就导致了指令假说的提出,该假说认为,有一种可变的抗体分子,该分子在抗原的作用下,能形成一个与抗原互补的结合部位。随着50年代到60年代分子生物学领域的飞速发展,指令假说(instructive hypothesis)显得越来越站不住脚,反过来又有更多的人倾向选择性假说。Jerne和Burnet提出了克隆选择的理论,他们认为,每个淋巴细胞只产生一种免疫球蛋白,抗原选择并刺激携带这种免疫球蛋白的细胞。

但是抗体多样性问题仍有待解决。一种解决方法

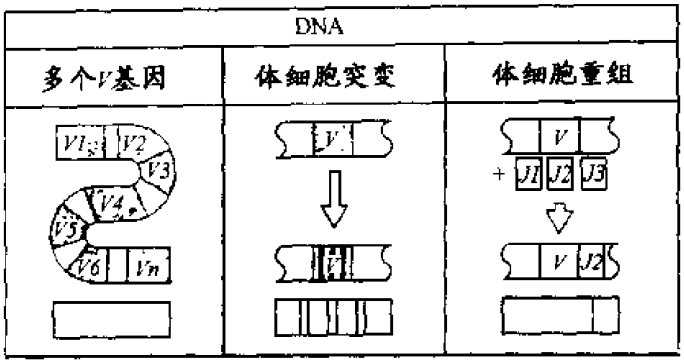


图 12-2 抗体多样性的产生

免疫系统在免疫球蛋白的 H 和 L 链中产生不同 V 区的三种机制：(1)多个基因 有大量不同的基因(*V*1～*Vn*)，每一个基因编码一个 V 区的结构域；(2)体细胞突变 在 B 细胞个体发育过程中，一个原始的 V 基因发生突变，因此在不同的 B 细胞克隆中产生不同的基因；(3)体细胞的重组 在 B 细胞个体发育过程中，许多基因片段(*J*1～*Jn*)能与 V 区基因发生重组。因此重组基因编码的蛋白质中所含的成分是由不同的基因片段编码。

在哺乳动物体内，抗体多样性的产生涉及到上述所有三种机制。

是，假定存在针对每种抗体特异性的不同基因。这就立即产生一个问题，观察一个轻链的结构，其氨基酸序列一半是可变的，另一半是恒定的。相似的情况也见于重链，重链的四分之一是可变的，其余部分是恒定的。如果存在这样多的基因的话，恒定区序列的恒定性是如何维持的呢？Dreyer 和 Bennett 提出一个解决问题的方法，他们认为，链的恒定区和可变区分别由不同的基因编码。一个或几个基因编码恒定区(constant region，C)。可变区(variable region，V)则由许多基因编码。在这种情况下，该理论只需要解释多个可变区的问题。体细胞突变的观点，为多样性问题提供了一种解释。该观点认为，在性细胞中有相对较少的基因，但在人的一生中，能产生许多突变的基因。同时他们还认为，许多基因片段可以重组，产生一个完整的 *V* 基因。这样就有三种产生多样性的学说：

(1)种系学说 该学说认为在性细胞内有多个 V 区基因。

(2)体细胞突变学说 该学说认为在个体发育过程中体细胞发生突变获得大量突变基因。

(3)基因重组学说 体细胞基因片段间的重组形成 V 区基因。

现在知道，哺乳动物利用上述所有三种机制产生多样性(图 12-2)。

有趣的是，鲨鱼依赖体内大量抗体基因，而不发生体细胞基因重组；鸟类有较少数目的抗体基因，但这些基因发生很高水平的基因转换。如后边将要讨论的，通过轻链和重链的随机配对，以及 *V*、*D*、*J* 重组过程中一些微细的变化，可引入更多的多样性。

二、免疫球蛋白的变异性

免疫球蛋白由重链和轻链组成，轻链有两种：κ 或 λ。实际上，任意一个轻链能与任意一个重链相结合，所以抗原结合部位的可能数目是重链和轻链数目的乘积。免疫球蛋白的多样性部分来自这些不同多肽链之间的结合。例如，如果有 10^4 不同的轻链，每一个能和 10^4 不同的重链中任何一个相结合，这样在理论上就产生 10^8 不同的抗体特异性。因

为它们分别由不同的染色体编码，所以每条肽链多样性产生的机制是不同的(表 12－1)。

表 12－1　MHC 和抗原受体基因的染色体定位

肽	鼠	人
IgH	12	14
λ	16	22
κ	6	2
TCRα	14	14
TCRβ	6	7
TCRγ	13	7
TCRδ	14	14
MHC	17	6
β_2-微球蛋白	2	15

表中数目是指人和鼠的不同多肽链的基因在染色体上的定位。

注意：所有的基因座是完全不同的，只有一个 T 细胞受体 TCRδ 链例外，它位于 TCRα 的基因座之内。

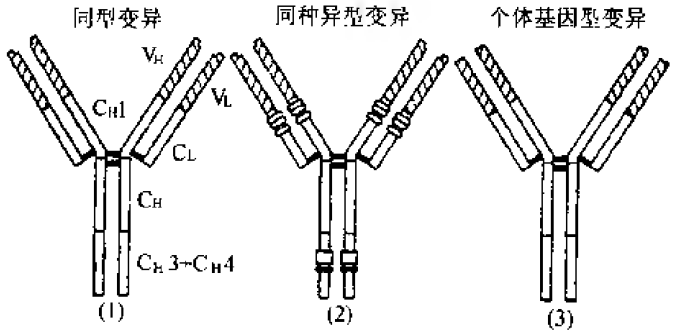


图 12　3　免疫球蛋白结构的变异性

所有的免疫球蛋白皆有基本的四链结构，免疫球蛋白的变异性有三种类型：(1)同种型变异　是指一个物种所有个体的性细胞中的重链(μ、δ、γ、ε、α)、轻链(Igκ、λ)和 V 区结构框架(亚群)的差异；(2)同种异型变异　是指种内等位基因的变异性；(3)个体基因型变异　是指抗原结合部位，特别是高变区片段的多样性。

免疫球蛋白的多样性也源于分子内部许多部位的差异(图 12－3)，这里我们首先讨论与抗原结合部位多样性有关的个体基因型的差异。

Kabat 等分析了许多重链和轻链可变区的氨基酸序列。为了得到相同抗体，他们对杂交瘤(一种产生单克隆抗体的 B 细胞杂交瘤)进行研究，结果发现，λ 轻链的变异性集中在三个高变区，其两侧是相对恒定的结构框架。后来证明，这些高变区正是与抗原接触的部位，被称为互补决定区(complementarity-determining region，CDR)。

开始时，他们以鼠作为研究对象，鼠含有 λ 轻链抗体的数目低于 5%，因此多样性相对少些(这是因为鼠含有很少数目的 $V\lambda$ 基因)。他们分析了 19 个轻链的序列，其中有 12 个是相同的，另外的 7 个不同，并且它们之间只有少数几个残基有差异(图 12－4)。

以 CDR 两侧的变异性为背景，重链的变异同样集中在三个高变区(图 12－5)。重链的结构框架(在 CDR 之间的序列)，按着它们序列的相同性或相似性分成不同的组(图 12－6)。

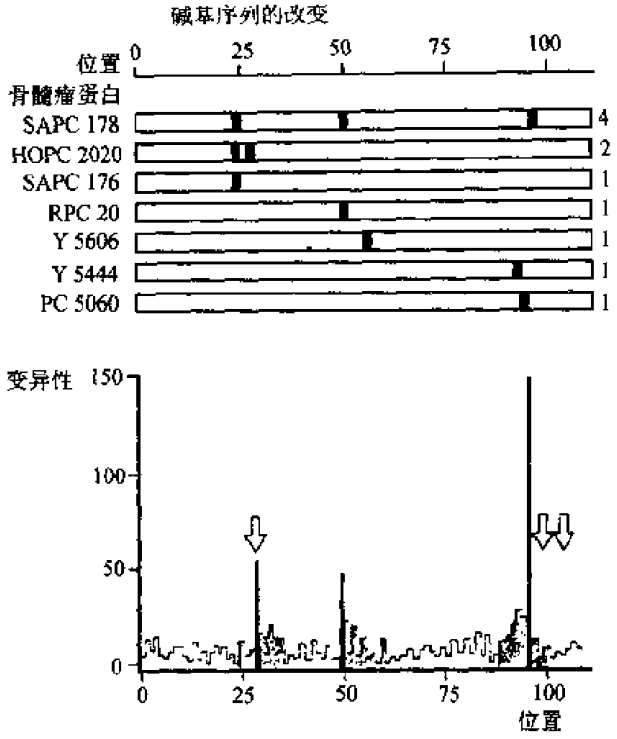


图 12-4 λ轻链的变异性

图示 7 个鼠 λ1 骨髓瘤蛋白氨基酸序列，空白位置表示与原始序列(MOPC 104E)相同，黑色部分表示和原始序列不同。产生已知氨基酸结构改变所需的相应 DNA 序列中碱基数目的改变，在右侧标明。下面为轻链变异性的另一种表示方法，其纵座标由已知位置上不同的氨基酸数目除以最常见的氨基酸出现的频率而得出。箭头表示在某些位置上存在额外的氨基酸。

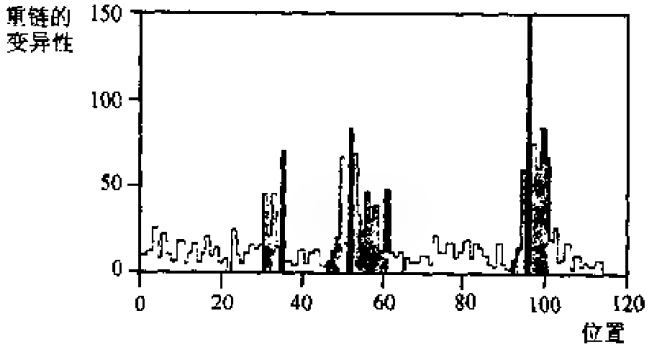


图 12-5 重链的变异性

Kabat 等指出重链的变异性集中在重链可变区的三个部位。

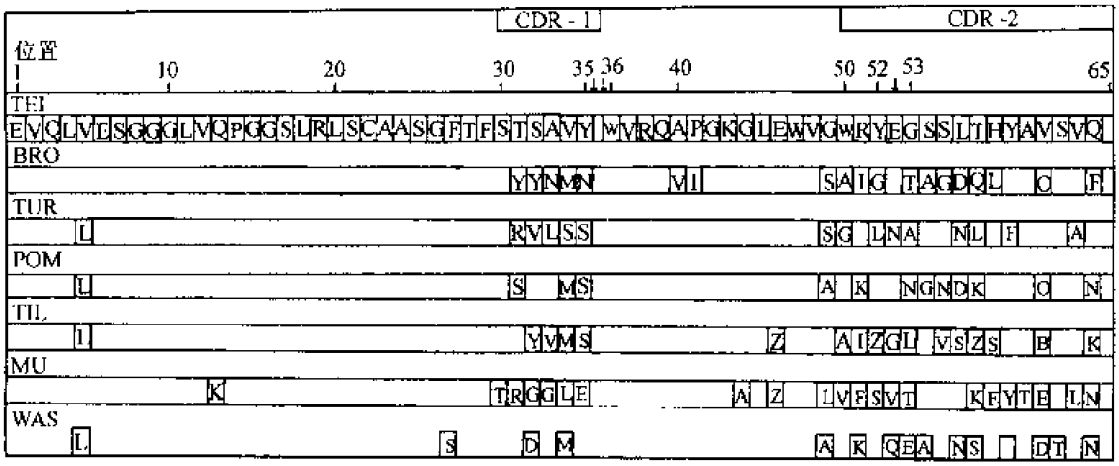


图 12-6　人类重链 V_HIII 组

图示 6 种属于 V_HIII 组的人类骨髓瘤重链 N 端 65 个氨基酸组成与原始序列 TEI 的比较。空白处表示氨基酸与 TEI 相同。在 TEI 相同位置上，不同的氨基酸用方框表示，并用氨基酸缩写字母标明氨基酸的种类，每种重链的不同氨基酸绝大部分出现在 CDR-1 和 CDR-2 区，第三个高变区在此图未表示。

第二节　免疫球蛋白基因的重组和体细胞突变[1,2,4,5~10,13]

一、轻链基因 *V* 和 *J* 片段的重组

70 年代 DNA 重组技术的出现，使分析编码抗体基因的愿望成为可能。因为鼠的 *λ1* 的多样性较少，所以工作首先在鼠的 *λ1* 系统进行（鼠有 4 个同型 λ 基因，依次排列在 λ 基因座，分别为 *λ1*～*λ4*）。现已发现恒定区和可变区是由两个不同的基因片段编码。在不产生抗体的细胞中，在染色体上，这些基因片段彼此远离。但在产生抗体的细胞中，它们靠的很近。然而，即使在完全成熟的分化的 B 细胞中，它们仍被大约 1 500 个碱基对隔离。在被重排的染色体上，*V* 和 *C* 片段之间，有一个较短的和 *V* 片段相连的 DNA 片段，被称为 *J* 片段（joining segment，*J*）（注意：*J* 片段不应当与在 IgM 和二聚体 IgA 中的 J 链相混淆）。

（1）λ 链系统的结构　　编码抗体轻链 V 区氨基酸残基 1～95 的是 *Vλ* 基因片段，由 *Jλ* 基因片段编码 *V* 区的其余部分（图 12-7）。

鼠类中（非人类），λ 轻链系统基因片段的总数目，与其他免疫球蛋白或 T 细胞受体基因座相比是有限的。已经发现有 3 个 *Vλ* 基因片段和 4 个 *Jλ-Cλ* 片段（图 12-7），其中 *J* 片段 *Jλ4* 是一个假基因。重组几乎总是在一个 *V-J-C* 片段区内进行，所以只有 4 种结合的可能性。它们分别是 *Vλ2-Jλ2*，*VλX-Jλ2*，*Vλ1-Jλ3* 和 *Vλ1-Jλ1*（见图 12-7）。每个 *V* 片段的前部有一段信号或前导序列编码 λ 链上的一个短的疏水氨基酸序列，该序列将 λ 链引导并定位于内质网。前导序列在内质网中被切除，然后抗体分子进入到细胞内分泌途径。

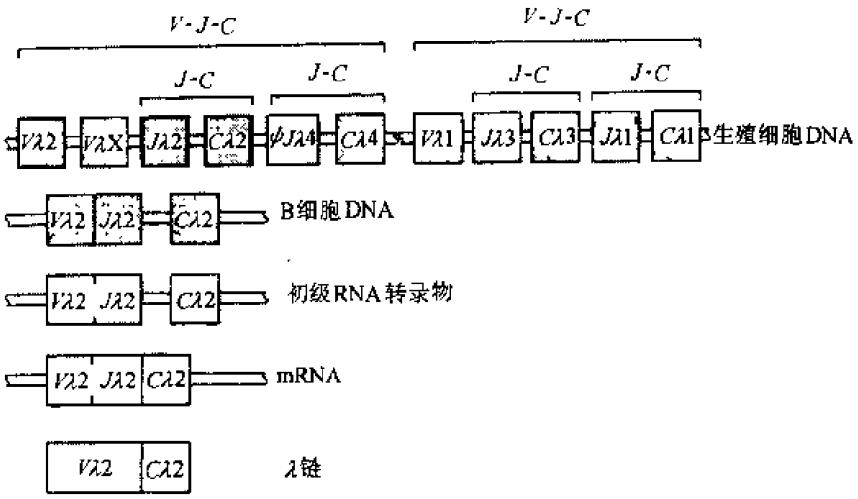


图 12-7 鼠λ链的形成

在B细胞分化的过程中，性细胞中的 Vλ 基因之一和相连的一个 J 基因片段重组，形成一个 V-J 重组体。重排的基因被转录成 RNA 的初级转录物(含内含子、外显子和一个 polyA)。经过剪接形成成熟的 mRNA，然后被翻译成蛋白质。

(2)κ链系统的结构　该系统有较多的 V 基因，因此即使只有一个恒定区基因，也有较高的多样性(图 12-8)。胚胎细胞或非淋巴样细胞内，大约有 350 个 Vκ 基因片段，

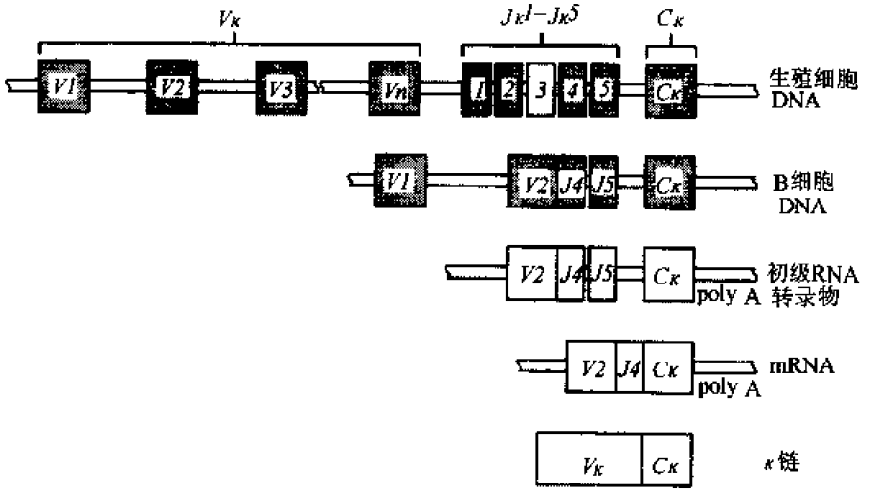


图 12-8 鼠κ链的形成

B细胞分化过程中，性细胞 DNA(V1～Vn)中的几个 Vκ 基因中的一个，与一个 Jκ 基因片段(Jκ1～Jκ5)发生重组。B细胞内这一段 DNA 片段，被转录成一个初级 RNA 转录物，含有一个由多余的 J 片段和内含子组成的长间隔序列。该转录物被剪切后，外显子部分相连，变成成熟的 mRNA，并通过核糖体翻译成 κ 链。注意，Jκ3 基因缺乏进行重组必需的碱基序列。因此，它实际上是一个内含子。该图所示，只是许多可能的重组类型中的一种。

在染色体上与 Cκ 基因有一定距离。鼠 Vκ 基因片段似乎被组成一定的系列，每个系列内含有 7 个基因。Vκ 基因和 C 基因之间，靠近 C 基因，是 5 个 Jκ 基因，其中之一是假基

因，从不表达。

在淋巴细胞分化时，DNA 重排，一个 *V* 基因与一个 *J* 基因相连，以这种方式产生的 κ 链可变区的可能数目大约是 1 400(350×4)。重排后，在 *J* 基因和 *C* 基因之间仍然有一个内含子。这一整段 DNA(由前导序列到 C 基因末端，包括内含子)被转录成不均一核 RNA(hnRNA)即未成熟的 mRNA，然后经过剪接过程除去内含子，成熟的 mRNA 经翻译产生蛋白质。

二、重链基因 *V*、*D* 和 *J* 片段的重组

重链可变区也是由 *V* 和 *J* 基因片段编码，附加的多样性由第三个 *D* 基因片段(diversity，D)提供(图 12-9)。*D* 基因片段的密码子的数目和碱基对的序列都是高度可变的。在葡聚糖的抗体中，这部分由 2 个氨基酸组成；磷酰胆碱的抗体中高达 8 个氨基酸被插入，而在抗果聚糖抗体中，该部分完全缺失。2 个以上的 *D* 片段可以相结合，形成一个扩展的 *D* 区。该 *D* 区可以被三种可能的可读框所阅读，不产生终止密码子，因此也增加了多样性。迄今，30 个性细胞的 *D* 片段，100～200 个不同 *VH* 片段，6 个功能性的 *J* 基因片段和 3 个假 *J* 基因片段，已经被鉴定。但是，研究表明，事实上可能有多于 1000 个鼠的 *VH* 基因片段。重链 *V*、*D* 和 *J* 片段重组结合点，对互补决定区 CDR-3 的多样性起很大作用，此区域构成了抗原结合部位的主要部分。某些系统，例如，抗葡聚糖抗体家族，抗体之间的差别几乎全部集中在该区。

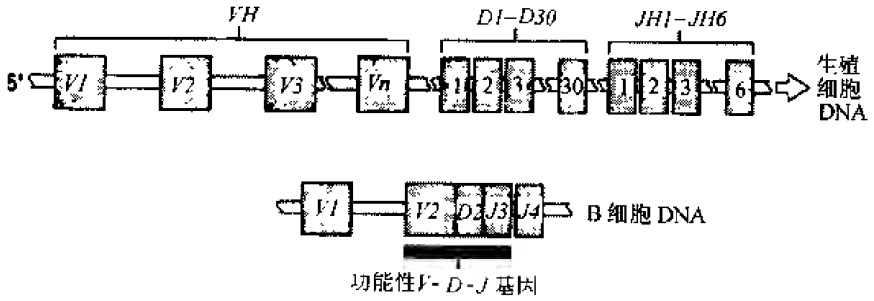


图 12－9　鼠的 *V-D-J* 基因重组

重链基因座的三个基因片段 *V-D-J* 结合产生编码 V_H结构域的外显子(*V-D-J* 基因)。几百个 *V* 基因之一和 30 个 *D* 片段之一及 6 个 *J* 片段之一重组，产生一个功能性的 *V-D-J* 基因。该重排图所示，只是数千种可能性之一。

三、*V*、*D* 和 *J* 基因两侧的序列指导基因片段的连接

基因片段的重组是产生轻链和重链可变区功能性基因的关键。发生重组的确切机制还不清楚，但已经发现一些特殊性的碱基序列似乎起连接信号的作用(图 12－10)。

在 *V* 和 *D* 基因片段的下游 3′端，有一个信号序列，它是由一个 CACAGTG 七聚体或相似的序列与一个非保守的间隔序列相连接，然后和一个 ACAAAAACC 九聚体或相似的序列相连。所有的性细胞中，紧接着 *D* 和 *J* 片段上游的 5′端，有一个相应的信号序列，它

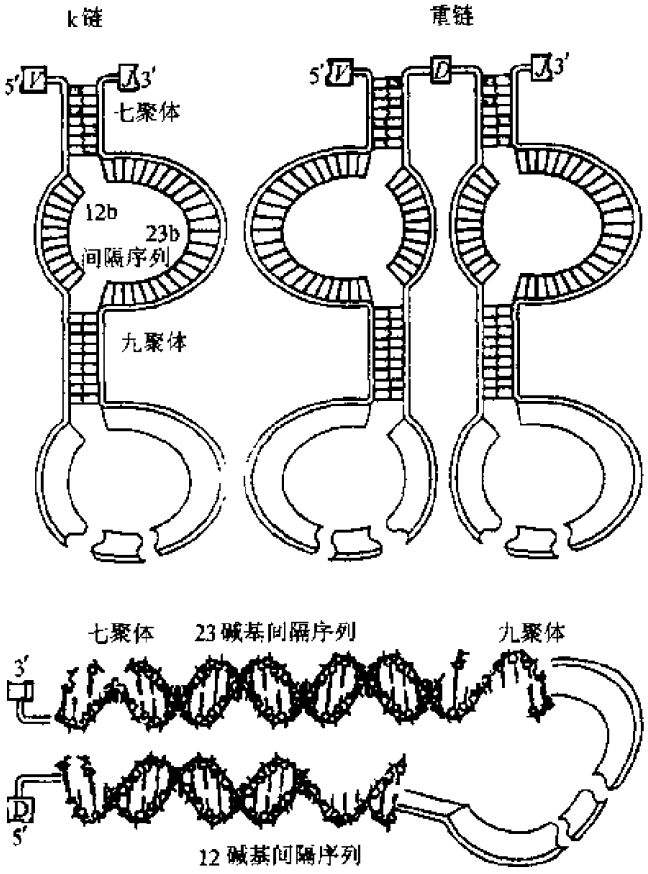


图 12-10 重组序列

研究认为，涉及 *V-J*（κ轻链，左上图）和 *V-D-J*（重链，右上图）剪接的重组过程，被 *V* 和 *D* 的 3′末端与 *J* 和 *D* 的 5′前端相匹配的内含子序列促进。这些序列之间的配对，使外显子相互接近。值得注意的是，每个碱基对序列可能有轻微的变化，但是七聚体－间隔序列－九聚体配对的模式不变，与 DNA 修复有关的酶影响连接过程。有迹象表明，当 V 区和 J 区 DNA 被引导靠近之后，发夹式的连接有时发生在编码重链的双链 DNA 断端之间（下图）。这表明，重组可能涉及双链 DNA 七聚体和九聚体区域的识别。

是由一个九聚体序列、一个间隔的非保守序列和一个七聚体序列所组成。*VL*、*VH* 或 *D* 片段的 3′端的七聚体及九聚体序列，和与它们重组的 *JL*、*D* 或 *JH* 片段 5′端的九聚体及七聚体是互补的。如果互补的七聚体和九聚体相配对，则产生两部分不配对的碱基，一个为 12 个碱基长，另一个为 22～24 个碱基长，相当于 DNA 的 1 个或 2 个螺旋周期。把基因片段连接在一起的重组酶能识别这一结构。

重组过程至少部分被两个重组活性基因 *RAG-1*、*RAG-2*（recombination activating genes，*RAG-1*、*RAG-2*）所控制。目前还没证实这些基因是否编码重组酶或一种调节重组的活性因子。但已证明在 *RAG-1* 和 *RAG-2* 基因敲除（knocked out）的鼠体内，由于不能产生 T 细胞受体和免疫球蛋白，因此体内缺乏成熟的 T 细胞和 B 细胞。研究表明，一种由骨髓间质细胞产生的细胞因子（cytokine）白细胞介素 7（interleukin IL 7），影响 *RAG-1* 和 *RAG-2* 表达水平。

四、*V*、*D* 和 *J* 基因片段的连接部位可以有轻度变化

重组发生位置的轻度差异，也能增加抗体的多样性。例如，κ 链的第 95 个残基通常由 *V* 基因片段最后的一个密码子编码，第 96 个残基常常由 Jκ 链的第一个三联体编码，但有时第 96 个氨基酸残基由组合型三联体编码，该三联体只含 Jκ 链第一个三联体的第二位和第三位或只是单独第三位碱基，其它碱基则来自与 *V* 基因相连的内含子的 3′端（图 12－11），这将导致氨基酸序列上的差异。很明显，为了产生一个功能性的轻链，必须保持一个正确的可读框，如果阅读框架发生错误，则产生一个无功能的抗体。

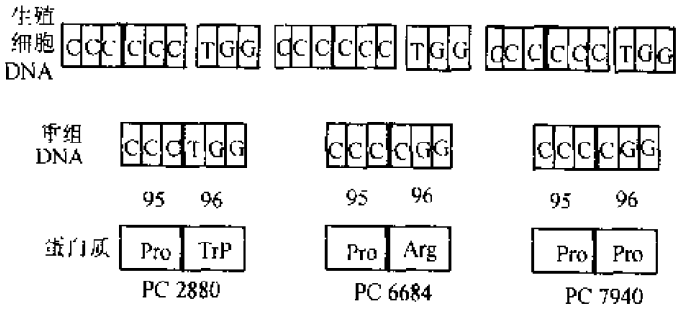


图 12－11　重组位置的差异产生轻链的多样性

在性细胞中相同的 *Vκ21* 和 *J1* 序列，通过不同组合，产生三种不同氨基酸序列的蛋白质 PC 2880、PC 6684 和 PC 7940　PC 2880 的 95 位和 96 位是脯氨酸和色氨酸，是由末端的 CCC 密码子重组产生。重组时 CCC 末端多带一个碱基，则产生脯氨酸和精氨酸（PC 6684），如重组时多带两个碱基，则产生脯氨酸和脯氨酸（PC 7940）。

一些额外的碱基也可以加入到编码片段的末端，有证据表明，*RAG*1、2 能识别带有 12/23 间隔的七/九聚体信号并在一条肽链适当的连接部位产生一个切口，切口处产生 3′－OH 和 5′－P末端，然后，游离的 3′－OH 末端攻击在该双螺旋另一条链相应位置上的磷酸酯键，这样，在编码片段末端的一条链的 3′端与另一条链的 5′端共价相连，产生一个发夹结构。在切口的另一侧留下一个平头末端。

编码区片段末端的发夹为下一阶段的反应提供了底物，在靠近发夹部位一个单链被切断，产生一个单链突出的末端，然后经互补合成产生一个延长的编码链末端。这种反应被认为是在编码片段末端引入了 P 核苷酸；它们由几个额外的碱基对组成，目前还不知道催化该反应的酶是什么。一些额外的碱基也可以随机地加入到编码区末端，它们称为 N 核苷酸，它们是在末端脱氧核苷酸转移酶（已知是淋巴细胞内的一种活性成分，该酶不需要模板）作用下，使少量核苷酸加入到连接的片段之间（图 12－12）。

在重链，*D* 和 *J* 基因片段之间不同位置的连接也可发生，有时差异可多达 10 个核苷酸（图 12－13）。而且，通过末端脱氧核苷酸转移酶的作用，少数核苷酸可以被插入到 *D* 和 *JH* 基因片段之间。这些 N 核苷酸的加入，被称为 N 区（non-template encoded，N）多样性。鼠随着年龄的增加，末端脱氧核苷酸转移酶的活性也随之增加，因此成年鼠含较长的 N 片段。这样，*D* 区的重组多样性是如此之明显，以致于没有可识别的残留 *D* 基因片段。

这些不同的机制一起确保编码片段的连接可以产生一个序列，该序列与 *V*、*D* 和 *J* 区编码片段末端直接连接所产生的序列不同。连接时序列的改变使该部位的氨基酸序列

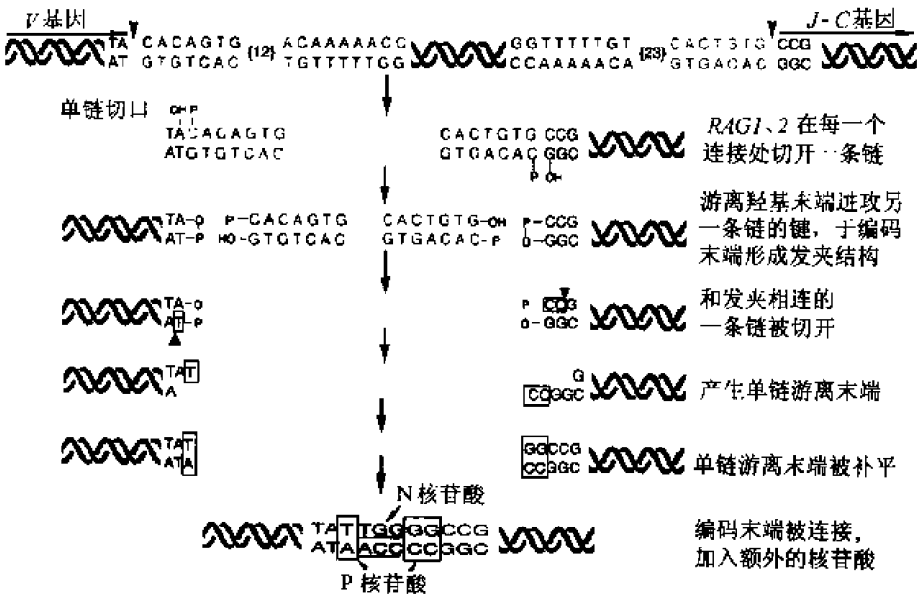


图 12－12　编码片段末端连接时 P 和 N 核苷酸的掺入

(1)*RAG*1:2 在每一个连接部位切断一条链；(2)游离的 3′－OH 攻击另一个链上的磷酸酯键，在编码片段末端产生发夹结构；切口的另一侧为平头末端；(3)与发夹相连的一条链被切开；(4)被切开的链产生一个单链突出的游离末端；(5)单链末端被补平；(6)当编码片段末端相连接时，额外的核苷酸被加入。

产生较大差异。

有严重免疫联合缺陷病(severe combined immune deficiency，SCID)的鼠，由于 *V*-(*D*)-*J* 重组的缺陷，不能产生有功能 T 或 B 细胞。

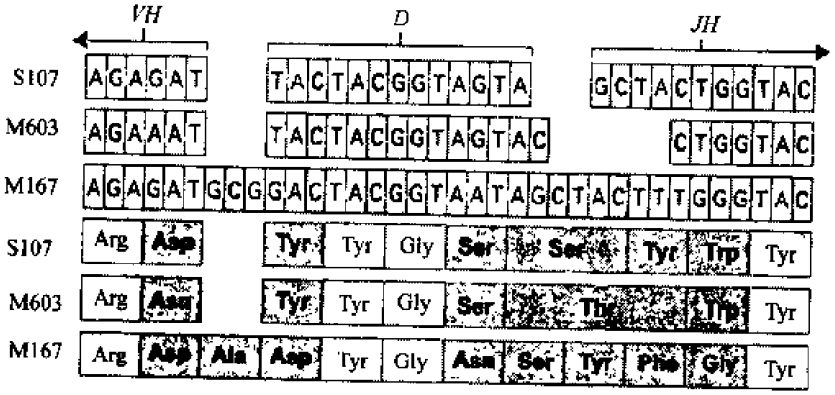


图 12－13　重组位置的差异产生重链的多样性

图示抗磷酰胆碱抗体的三个重链 DNA 序列(上)和氨基酸序列(下)。在性细胞中，*V*，*D* 和 *J* 区不同部位之间的重组，导致氨基酸序列的差异。在某些情况下，(如 M167)，有多余的插入密码子，然而，因为插入的碱基的个数为 3 的倍数，因此整个可读框不变。

五、体细胞突变

免疫球蛋白的轻链和重链在抗原刺激后结构发生改变。一个人的生命进程中，体细胞发生突变，并以此增加抗体多样性的观点已争论了许多年。像前面所介绍的(参见图12－4)，鼠λ1轻链大部分序列都是一样的，只是由于互补决定区的一些差异，而形成8个不同的序列。然而，在单倍体基因组中，只有一个 *VλI* 基因片段，即为原型序列，这样，所有序列的改变是通过体细胞单个碱基的突变引起。体细胞突变也见于κ轻链和重链。

进一步的研究支持这种观点，在抗磷酰胆碱抗体基因家族的研究中，19个抗体的 *VH* 基因片段被测序，其中有10个和原型序列相同，其他9个序列存在1～8个残基的差异。性细胞基因组只含有原型序列。这表明其余的序列一定是由体细胞突变所产生(图12－14)。令人惊奇的是，所有的突变只见于IgA和IgG。这表明突变和免疫球蛋白类型的转换有关。那些能与相应抗原更好地结合的体细胞突变体，可能被选择。像在抗磷酰胆碱抗体的变异体中所见到的，和性细胞编码的抗体相比，变异体有较高的亲和力。

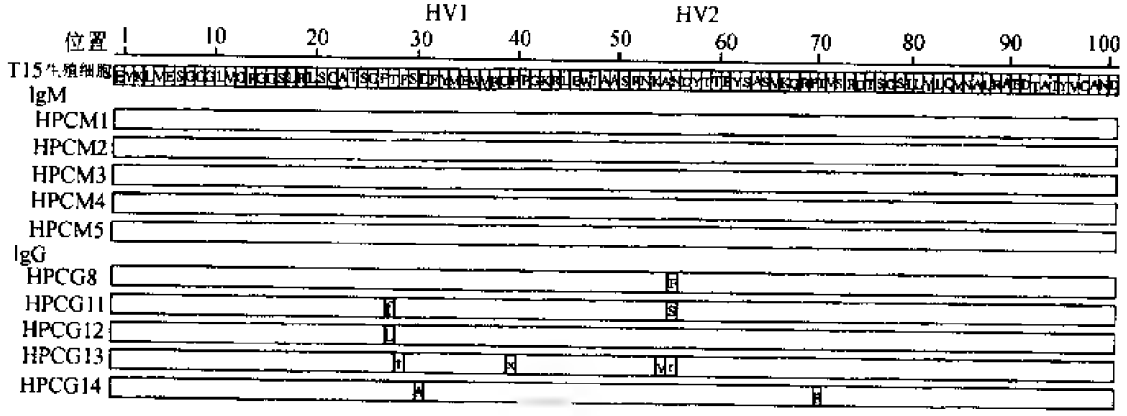


图12－14　体细胞突变

5个IgM和5个IgG抗磷酰胆碱杂交瘤抗体的V_H区的氨基酸序列，与T15性细胞的原型序列相比较，发生突变的氨基酸位点用方框标出，高变区用HV1和HV2标明。图中可见，突变只见于IgG分子，并且突变既发生在高变区也发生在框架结构中。

有迹象表明，编码可变区的DNA片段对突变特别敏感，例如，从对两个抗磷酰胆碱抗体的核苷酸序列的检测中发现，与性细胞相比，有较多的突变发生(M167蛋白有3.8%的碱基突变)。这些突变见于 *V* 区的外显子和内含子，但不见于连接序列，这表明整个 *V* 区特别容易突变(图12－15)。

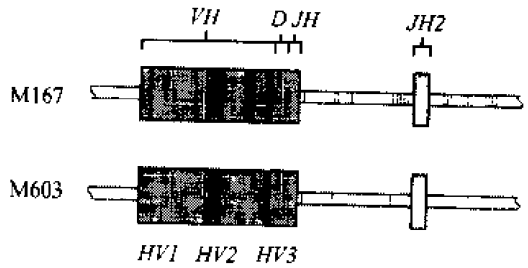


图12－15　两个 *VH T15* 基因DNA的突变

两个 *T15* 基因型的抗磷酰胆碱抗体的DNA，黑线标明和性细胞相比，基因组发生突变的位置。两个基因的外显子和内含子皆有大量突变，特别是在第二个高变区突变较多，在编码恒定区的基因中，没有检测到突变。

综上所述，抗体的多样性是通过几种不同的形式产生的。首先是多个 *V* 基因与 *J* 和 *D* 片段重组，重组部位的不同，进

一步增加了多样性。值得注意的是，第一个和第二个高变区的结构完全由性细胞编码，而第三个互补决定区的差异性，主要由重组产生。另外，整个可变区内发生的点突变，也能在抗体特异性上产生细微的差异。实际上，任何一个轻链可以与任意一个重链配对，重链和轻链的随机结合大大地增加了抗体的多样性。体细胞突变可能低于总的序列变异性的5%，但B细胞表达的 *VH* 基因发生突变高达90%。

产生抗体多样性的5种机制：

1）性细胞内含多个 *V* 基因；

2）*V-J* 和 *V-D-J* 的重组；

3）重组部位的差异性；

4）体细胞突变；

5）重链和轻链的随机组合。

由于每种机制皆能与其他机制相伴随而发生，因此，免疫球蛋白产生的每个阶段，其多样性皆以乘积的方式增加。

第三节　免疫球蛋白重链恒定区基因[1,4,12~14]

所有种类的免疫球蛋白都用同一系列的可变区基因。当一个成熟的抗体形成细胞转变抗体类型时，只有重链恒定区发生变化。此观点已经通过对双骨髓瘤（double myelomas，同时在血清中存在二种单克隆抗体）的分析得到证实。对一个患多发性骨髓瘤病人的IgM和IgG抗体的研究表明，二种抗体的轻链及 V_H 区是相同的，只是恒定区分别由 μ 和 γ 构成。在一种淋巴细胞的表面，经常可以发现同时存在IgM和IgD两种抗体，虽然抗体的种类不同，但抗原结合的特异性是一样的。

所有的恒定区基因都位于 *J* 基因片段的下游，鼠的IgM、IgE和IgA每型各有一个恒定区基因（μ、ε 和 α），4种不同的IgG，每种各有一个 $\gamma 1$、$\gamma 2a$、$\gamma 2b$ 和 $\gamma 3$）（图12-16）。

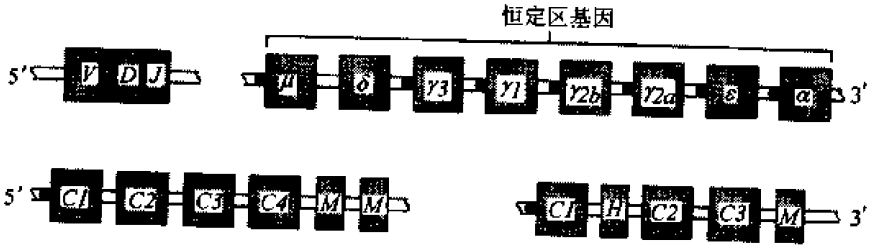


图12-16　鼠恒定区基因

鼠的恒定区基因位于 *V-D-J* 重组片段下游6.5kb处，每个 *C* 基因（*Cδ* 除外）皆有一个或多个转换序列位于它的起始部（图中黑圈处），它们与 μ 基因5′端的转换序列相对应。这样任何一个 *C* 基因片段都能与 *V-D-J* 基因发生重组。δ 基因似乎与 μ 基因共用同一转换序列，但 μ 基因的转录体在RNA剪接时被去除，产生IgD抗体。*C* 基因（μ 和 $\gamma 2a$ 基因被展开于图的下部）每个结构域（*C1*、*C2* 等）的外显子被内含子分开。γ 基因还有一个编码铰合区的外显子 *H* 基因（hinge，*H*），所有基因都有一个或多个参与编码膜结合免疫球蛋白M（membrane-bound immunoglobulin，M）的外显子。

人类恒定区的基因是较复杂的，有迹象表明，这个区的一部分已经发生了基因的复制

和分化。基因的顺序是 μ、δ、($\gamma 3$、$\gamma 1$、$\varepsilon 1$、$\alpha 1$)、γ、($\gamma 2$、$\gamma 4$、ε、$\alpha 2$),括号内的两组基因为发生再复制的可能区域。基因 $\varepsilon 1$ 和 γ 是假基因,不能表达。μ 基因的上游是一个转换序列 S(switch sequence,S),每个其他恒定区基因的上游皆有此转换序列(图 12 17)。

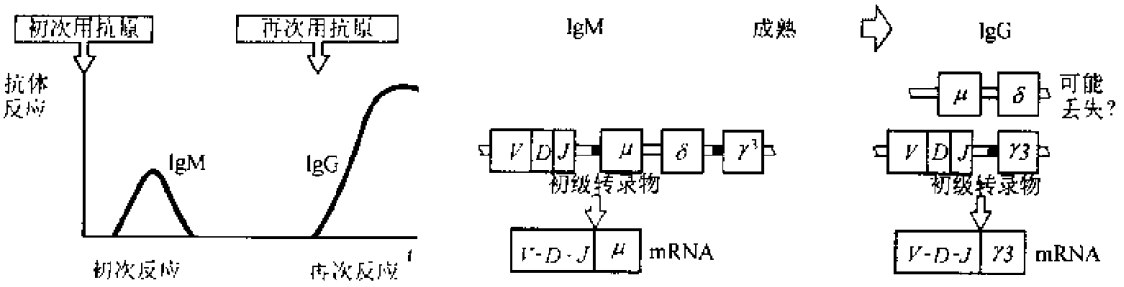


图 12-17 免疫反应的成熟过程和抗体类型的转变

第一次抗原注射后,引发一个抗体反应,并且产生的抗体主要为 IgM;第二次抗原注射后,主要产生 IgG 抗体。抗体类型转变的机制见右图。在第一次反应中,*V-D-J* 和 μ 基因一起被转录,通过剪切除去内含子后,产生分泌型 IgM。

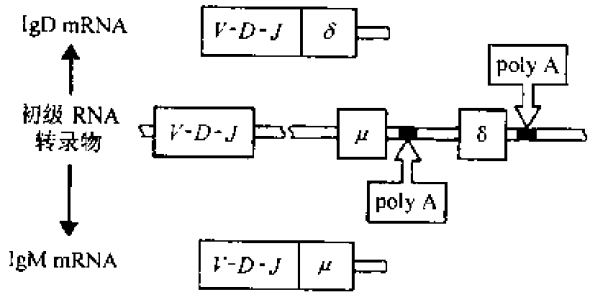


图 12-18 通过 RNA 的差异剪接产生同型转换

在一个淋巴细胞内,一个长的 RNA 初级转录体可产生两种以上同型抗体。图示一个含有 μ 和 δ 基因的转录物。因为多聚腺嘌呤可出现在不同部位,因此导致产生不同类型的剪接,产生 IgD 的 mRNA(上)或 IgM 的 mRNA(下)。甚至在 μ 和 δ 区域内仍有多聚腺嘌呤位点,这些位点能决定被翻译的免疫球蛋白是分泌型的还是结合膜型的。

免疫成熟过程中(涉及到 T 细胞的辅助和对应于 *V-D-J* 基因片段的一个可能的突变机制的激活),另一个 *C* 基因(此图为 $C\gamma^3$)被带到 *V-D-J* 基因片段附近,在 μ 基因的转换区(用黑圈表示)与其发生交换,μ 和 δ 基因可能丢失了。经转录和剪接后,产生编码 IgG 重链的 mRNA。

一、免疫球蛋白类型的转换

在免疫反应成熟的过程中,抗体类型的转换是很重要的,并且可能伴有(同时或先发生)体细胞突变。开始时,一个完整的 DNA 片段,包括重组的 *VH* 区到 δ 和 μ 恒定区被转录,通过差异剪接产生两种mRNA分子,此两种 mRNA 含有相同的 *VH* 区,但恒定区不同,一种是 μ,一种是 δ。目前认为,有时一段较长的 DNA 片段被转录,通过差异剪接产生不同的 mRNA,并通过翻译产生不同的免疫球蛋白,它们含有同样的 *VH* 区(图 12-18),但恒定区不同。在同时产生 IgM 和 IgE 的细胞中,已观察到这种情况。

更常见的是,抗体类别的转换,可能被 *S* 重组部位之间的重组所介导,通过环化部分 DNA 的去除,导致另一个 *C* 区靠近 *VDJ* 基因。另外,该过程可能和染色体间的交换有关(图 12-19)。

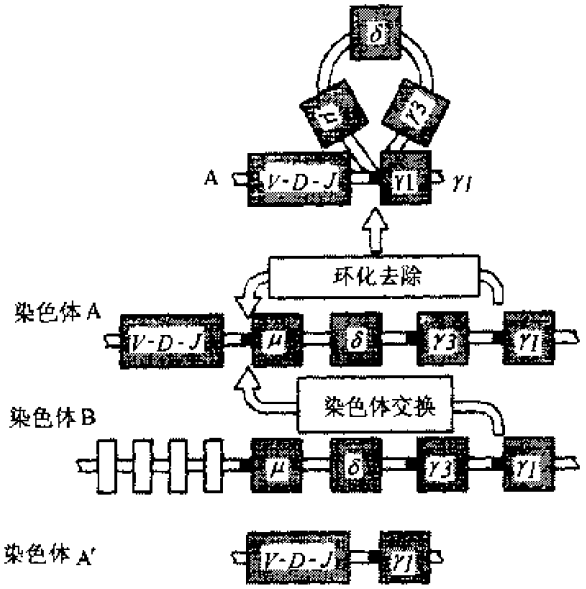


图 12-19 基因的丢失

A和B是含有免疫球蛋白基因染色体的染色单体。A含有重排的 *V-D-J* 片段。成环去除假说认为，*C* 基因的一部分（μ，δ 和 $\gamma 3$）形成一个突出的环并被去除。而染色体交换假说认为，转换序列的相似性，使父系和母系染色体可发生不等的体细胞重组。这样A染色体和未发生重排的B染色体一部分发生重组，并由于基因片段的丢失，产生了 IgM 到 IgG 的转换，表示为 A′。"丢失"的 *C* 基因片段则在另一个无功能的染色体 B′上（该图未示），此染色体含有几个 *C* 基因的两个拷贝。

二、膜型和分泌型免疫球蛋白的产生

膜型免疫球蛋白（抗原受体），除了在每条重链 C 末端有一条额外的氨基酸链之外，与分泌型免疫球蛋白（抗体）是一样的。因此膜免疫球蛋白略大于分泌型免疫球蛋白。它们多余的氨基酸部分穿过细胞膜并使整个分子嵌于脂质双层膜上。例如，在膜中的 IgM 部分是疏水（亲脂）氨基酸部分，位于膜两侧的是亲水氨基酸部分（图 12-20）。疏水氨基酸残基部分被认为在细胞内形成 α 螺旋。膜免疫球蛋白只以四链基本单位的形式存在，没有多聚体形成。

两种类型的免疫球蛋白是通过性细胞 C 区基因不同方式的转录而产生的（图 12-21）。有迹象表明，多聚腺嘌呤序列对决定产生哪种 RNA 转录体起重要作用。但目前还不清楚具体过程是如何控制的。

三、免疫球蛋白产生的调节

很明显，细胞调节免疫球蛋白生成的方式是很复杂的。第一步是重链 *D* 和 *J* 基因的重排，接着加入 *V* 区基因。μ 链刺激 κ 基因座，促进 *V-J* 重排，如果必要的话，μ 链也可刺激 λ 基因座并促进 *V-J* 重排。目前认为，该过程可以重复地发生在母系和父系的染色体上，直到一个有功能的重组基因形成，或遗传物质被耗尽并伴随细胞死亡。细胞的 *V* 区一旦形成，除非有任何体细胞突变，它们基本保持不变。但 *CH* 基因仍可发生转换而产生

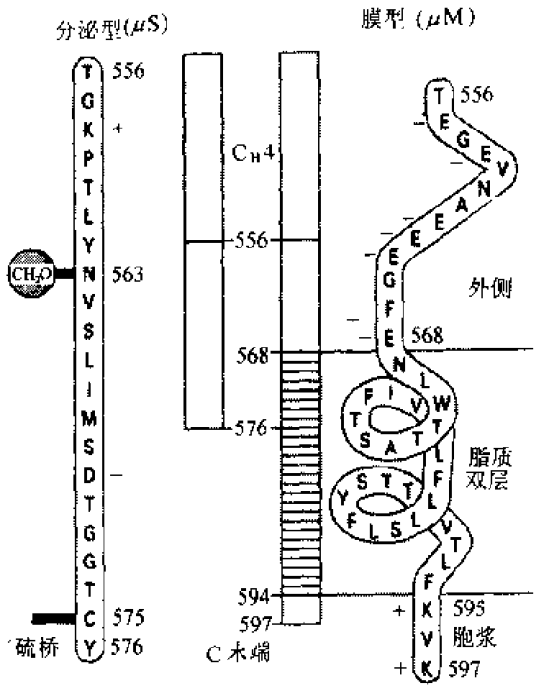


图 12－20　鼠的膜型和分泌型 IgM

分泌型和膜型的 IgM 的 C 末端氨基酸序列在第 556 个氨基酸之前是相同的。分泌型 IgM 另含 20 个氨基酸残基，第 563 位残基天冬酰胺与一个糖单元相连，而第 575 位残基则是一个与链间的双硫键形成有关的半胱氨酸。膜型 IgM 在第 556 位残基之后，另含 41 个氨基酸残基。在 568 和 595 位残基之间，是一个由 26 个疏水的残基组成的氨基酸链，疏水部分在膜内可形成两个周期 α 螺旋。该部分的两侧是带有电荷的残基，位于胞浆内的是一个短的，带正电荷部分。

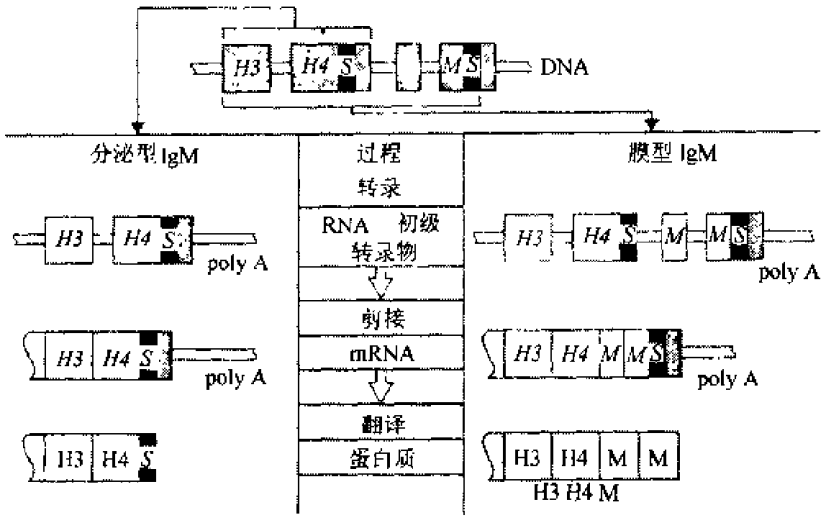


图 12－21　膜型和分泌型 IgM

图示编码 IgM 的 DNA 片段，其中有编码 μ3 和 μ4 结构域（*H3* 和 *H4*）和膜 IgM 穿膜部分的外显子（*M*）。在 *H4* 和第二个膜片段的末端，存在着 3′端非翻译序列（*S* 等于分泌型的 C 末端）。DNA 能以两种方式被转录，如果转录停止在 *H4* 之后，则带有一个 polyA 尾的转录物被剪接并产生分泌型 IgM 的 mRNA。如果转录继续进行下去，并包含膜片段，则剪切除去末端氨基酸的密码子和 *H4* 的终止信号，这样翻译则产生一个含有不同 C 末端的蛋白质。

各种不同类型的抗体，并在细胞被活化后，产生分泌型免疫球蛋白(图 12－22)。

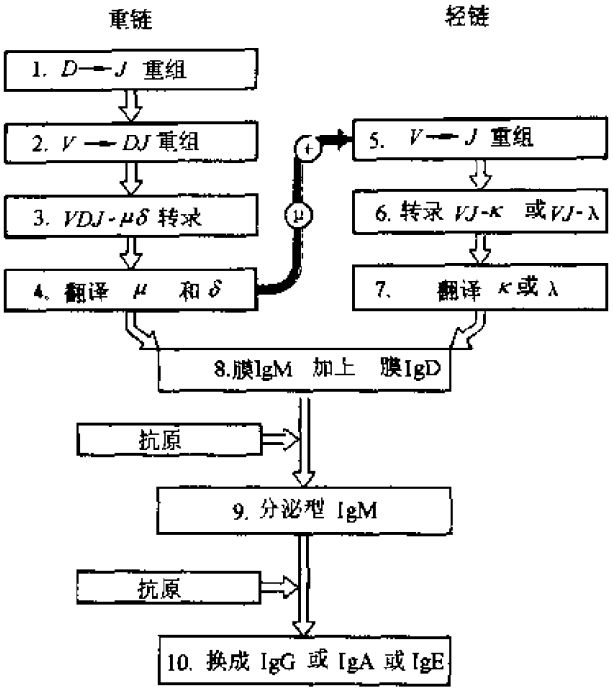


图 12－22 免疫球蛋白产生的总结图

图示免疫球蛋白产生的各阶段。未成熟的 B 细胞中重链 *D* 和 *J* 基因片段首先发生重组，然后这个重组体与一个 *V* 片段结合形成一个有功能的重链基因 *V-D-J*。该重组体与 μ 和 δ 基因一起被转录并翻译产生膜 μ 和 δ 链。μ 链诱导轻链基因座重组。在上述情况中，细胞尝试用父系和母系染色体产生有功能的链，直到成功或者未重排的基因被耗尽。不能产生有功能免疫球蛋白的克隆被淘汰。原始 B 细胞表达膜 IgM 和 IgD。在第一次抗原刺激后，细胞能产生分泌型 IgM，经过 T 依赖抗原刺激后，并在 T 细胞的辅助下，细胞能转变抗体的类型，产生 IgG、IgA 或 IgE，并且该阶段常常伴随有 *V* 区的体细胞突变。

合成免疫球蛋白之前，必须剪接掉其初级 RNA 转录体中的内含子。在每个内含子的起始和结尾部分皆存在特殊类型的 RNA 碱基序列，称为供体和受体结合点。目前认为，这些结合点彼此之间，以及与核蛋白在核内的相互作用，使内含子被去除并使外显子相连接形成 mRNA。当然，这些过程的正确进行并确保 mRNA 的可读框不发生改变，是十分重要的。成熟的 mRNA 被翻译，新生肽链与内质网相连，随后重链和轻链相连接(图 12－23)。膜型和分泌型免疫球蛋白通过不同的进程到达它们正确的位置，但其机制目前还不十分清楚。

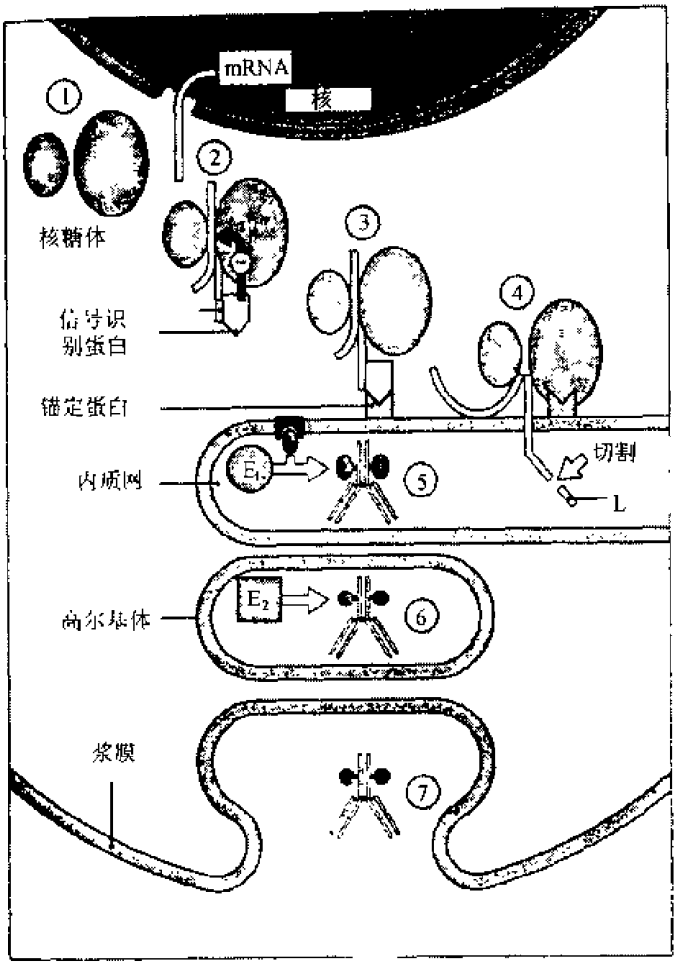


图 12-23　分泌型免疫球蛋白的产生

(1)分泌型免疫球蛋白重链的 mRNA 离开细胞核进入到胞浆;(2)前导序列 L(leader sequence,L)被翻译并与信号识别蛋白(signal recognition protein,SRP)相结合而阻止进一步的翻译;(3)SRP-核糖体复合物迁移到内质网。在那里,信号识别蛋白与内质网表面的锚定蛋白(docking protein)上的空位相结合;(4)在此处翻译继续进行,新合成的肽链穿膜进入到内质网;(5)切除前导序列,新合成的肽链和其它的 H 和 L 链结合,形成免疫球蛋白的亚单位;(6)酶(E_1)对免疫球蛋白进行加糖修饰,同时内质网被部分切离,形成一个转运载体;(7)在完整的分子通过胞饮的逆过程,被分泌到细胞外之前,酶(E_2)在高尔基氏体内对其糖基进行进一步的修饰。

第四节　T 细胞抗原受体基因[1,14,15]

一、T 细胞抗原受体基因

T 细胞抗原受体(TCR)的抗原-MHC 结合部分是由四组不同的基因编码的。α 和 β 基因群主要在外周的 T 细胞表达,而 γ 和 δ 基因主要在胸腺细胞亚群内表达,少数在外周 T 细胞内表达。这些链和 CD3 分子的 γ、δ、ε 和 ζ 链相连形成完整的 TCR 分子。

TCR 基因的重排和免疫球蛋白的重链的重排十分相似,有趣的是,TCR 的 δ 基因位于 α 基因座的内部,并带有自己的 D、J 和 C 基因片段(图 12-24)。

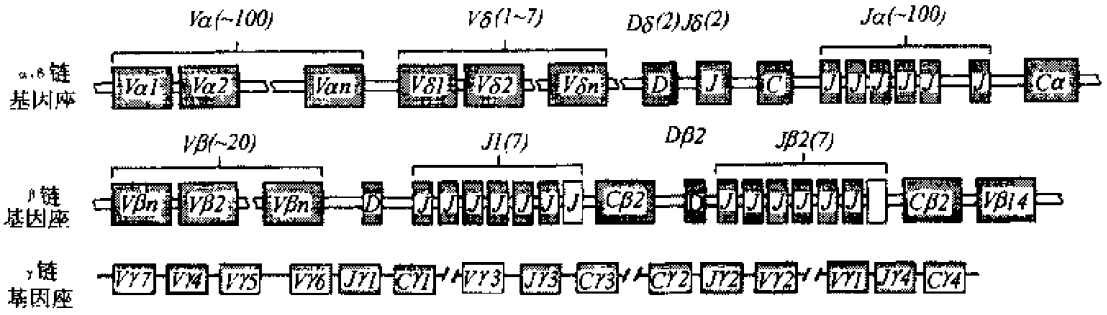


图 12－24　T细胞受体基因

T细胞受体基因重组的机制与免疫球蛋白重链基因重组机制相似。图示鼠的T细胞受体 αβ 和 γδ 多肽链的基因。δ链的基因座位于 α 基因座内。β链的部分基因发生了串联形复制。每组 *Jβ* 基因的最后一个和 *Vγ3* 基因是假基因。

TCR 基因的多样性是通过 *V*、*D* 和 *J* 基因片段重组产生，并且每个基因座之间有细微的差异。除了 δ 基因座位于 α 基因座 *V* 和 *J* 之间外，α 基因座结构表面上很简单。像在 κ 基因座一样，一个完整的可变区通过 *Vα* 和 *Jα* 片段的重排产生，由于存在大量的不寻常 *J* 基因片段，多样性因此大大增加了。

β 基因座包括两组 *D*、*J* 和 *C* 基因系列。除了一个 *Vβ14* 存在于该座位的 3′末端之外，其余的 *Vβ* 基因皆集中在一起。*Dβ*、*Jβ* 和 *Cβ* 的串联复制，一定发生在哺乳动物进化的早期，因为它们在鼠和人类中都存在。重组除以 *V-D-J* 的方式进行外，还可以 *V-J*、*V-D-D-J* 的方式进行，因此导致广泛的多样性产生。*D* 基因片段能以所有三种可读框被利用，这就进一步增加了β链的多样性。

鼠和人的 γ 链基因座的重排是相当不同的，鼠的该基因座与抗体轻链的基因座位明显相似，其中含有 4 个 *Cγ* 基因(包括一个假基因)。它们每一个都与 1 个 *J* 基因和 1 到 4 个 *Vγ* 基因相连，没有 *D* 基因。人有 8 个 *Vγ* 基因，接着是 3 个 *Jγ* 基因和第一个 *Cγ* 基因，在 *Cγ2* 前另有两个 *Jγ* 基因。*V* 和 *J* 基因片段间不同方式的连接，在连接点处残基的插入，对多样性产生起重要作用。

δ 基因座是在研究 *α* 基因座的过程中发现的。*δ* 基因座虽然相对简单，有 5 个 *Vδ*，2 个 *Dδ* 和 6 个 *Jδ* 基因，但据计算，通过连接方式的差异，附加残基的插入和 *D* 基因的所有 3 种可读框的应用，能产生 1 014 种不同的 δ 链。

TCR 基因重组的机制与 B 细胞的重组相类似，因为基因间含有类似的七聚体- 12 或 23 个碱基间隔 九聚体序列。在 B 细胞和 T 细胞中，有类似或相同的重组酶，操纵重组过程。

实验表明，被转染的 TCR *Dβ* 和 *Jγ* 基因，在 B 细胞内能进行适当的重排。特别应当指出的是 N 区的产生，大大地增加了多样性。

虽然体细胞突变是产生免疫球蛋白多样性的一种重要的方式，但 TCR 基因并不产生此种突变。这大概与维持免疫耐受性和 T 细胞对 MHC 分子的识别有关。

二、TCR 多样性的程度

多样性依赖于 V、D 和 J 的重组，N 区的多样性，连接部位的差异及多个 D 区的存在。性细胞内 V 区末端和 J 区之间的距离是可变的，在 $V\beta$ 此距离大约为 6～15 个氨基酸残基的编码序列，在 $V\alpha$ 为 3～7 个氨基酸残基的编码序列。Hunkapiller 等估计可能存在大约 4.4×10^{13}不同类型的 $V\beta$ 和 6.5×10^{12}不同类型的 $V\alpha$。他们估计，如果编码有功能的蛋白质的序列只占总数的 1%，则仍可产生 2.9×10^{22}个受体。假设其中 99% 的受体因编码自身抗原或含其他缺陷而被淘汰，则仍可产生 2.9×10^{20}种鼠的 TCR。鼠的一生中只有少于 10^9胸腺细胞离开胸腺，这就产生一个问题，受体的形成含有多大的随机性？同样的问题对免疫球蛋白也存在，可能产生的各种抗体的总数远远多于生物体一生中产生的 B 细胞数。

（张玉静）

参 考 文 献

[1] Roitt I et al. Immunology. London: Mosby, 1996.6:1
[2] Hunkapiller T et al. Adv Immunol, 1990.44:1
[3] Owen MJ et al. Immune Recognition, Oxford: IRL Press, 1988.
[4] Pascual V et al. Adv Immunol, 1991.49:1
[5] Lewin B. GENES, Oxford University Press, 1997.989
[6] Lewis S M. Adv Immunol, 1994.56:27
[7] Blackwell TK et al. Ann Rev Genet, 1989.23:605
[8] Brack C et al. Cell, 1978.15:1
[9] Chang B. Immunology Today, 1994.15:367
[10] Appasamy PM et al. J Exp Med, 1993.178:2201
[11] Haas W et al. Ann Rev Immunol, 1993 11:637
[12] Harriman W et al. Ann Rev Immunol, 1993.11:38
[13] Austen BM et al Oxford: IRL Press, 1991.
[14] Davis MM et al. Nature, 1988.334:395
[15] Leiden JM. Ann Rev Immunol, 1993.11:539

第十三章　DNA 损伤、修复和基因突变

第一节　DNA 损伤[1~5,27]

所谓 DNA 损伤即是指在生物体生命过程中 DNA 双螺旋结构发生的任何改变。大体上可分为两类：单个碱基改变和结构扭曲。前者仅影响 DNA 序列而不影响整体结构，当 DNA 双螺旋被分开时不影响转录或复制，而是通过序列改变作用于其子代，改变子代的遗传信息；后者则对复制或转录产生物理性的损害，如因紫外线辐射所致的结构扭曲，引起两个邻近 T 碱基之间产生共价键，形成链内 T 二聚体，或者某些原因形成单链缺口、凸起。引起 DNA 损伤的因素很多，有来自细胞内部的各种代谢产物和外界的物理、化学因素，还包括 DNA 分子本身在复制等过程中发生的自发性改变。

一、自发性损伤

1. 复制中的损伤

复制中的损伤指复制过程中碱基配对时发生的误差再经过 DNA 聚合酶“校读”和单链结合蛋白等综合因素作用下仍然存在的 DNA 损伤。如在大肠杆菌中，DNA 复制时发生的碱基配对的误配率为 $10^{-1}\sim10^{-2}$，而经过 DNA 聚合酶的作用之后校正为 $10^{-5}\sim10^{-6}$；再经过 DNA 结合蛋白质的校正作用和其他因素作用下误配率降至复制完成后的 10^{-10}，影响复制过程中的诸因素中如果某一环节出现问题，则误配率会增高，尤其是 DNA 聚合酶本身的功能和底物的改变，二价阳离子的改变等。

2. 碱基的自发性化学改变

这类损伤包括互变异构、碱基的脱氨基作用、自发的脱嘌呤和脱嘧啶，以及细胞代谢产物对 DNA 的损伤。

互变异构指 DNA 分子中的 4 种碱基自发地使氢原子改变位置，产生互变异构体，进一步地使碱基的配对形式发生改变，这样在复制后的子链上就可能出现错误。如腺嘌呤的互变异构体(A′)可以与胞嘧啶(C)配对，胸腺嘧啶(G)的互变异构体与鸟嘌呤配对，当 DNA 复制时，如果模板链上存在这些互变异构体，在子链上就可能发生错误，形成损伤。

碱基的脱氨基作用是指胞嘧啶(C)，腺嘌呤(A)和鸟嘌呤(G)分子结构中都含有环外氨基，氨基有时会自发脱落，结果 C 变为尿嘧啶(U)据估计，C 变成 U 的频率约为 8(细胞·小时)，A 变为次黄嘌呤(Hx)，G 变为黄嘌呤(X)，这些碱基的衍生物配对特性与原来的碱基不同，而是 U:A，Hx:C，X:C，当 DNA 复制时，会在子链中产生错误而导致 DNA 损伤。在生理条件下，DNA 可以通过自发水解使嘌呤碱和嘧啶碱从 DNA 的磷酸脱氧核糖骨架上脱落下来。据估计，一个哺乳动物细胞在 37℃ 条件下，20 小时内通过自发水解可以从 DNA 链上脱落约 1 000 个嘌呤碱和 500 个嘧啶碱。在一个长寿命的哺乳动物细胞

(如人的神经细胞)的整个生活周期中自发脱落嘌呤数约为 10^8 个嘌呤碱,占细胞 DNA 中总嘌呤数的 30%。每个细胞每小时脱去的嘌呤碱和嘧啶碱数分别平均为 580 个和 29 个。

O^{2-} 和 H_2O_2 是细胞呼吸的副产物,它们非常活跃,能造成 DNA 氧化损伤,产生胸腺嘧啶乙二醇、胸苷乙二醇和羟甲基尿嘧啶等,这些损伤基本上能被修复,损伤的碱基通过尿排出。此外还有葡萄糖和碱基氧化产物 6-磷酸葡萄糖,这些糖分子能和 DNA 反应,产生明显的结构和生物学改变。

二、物理因素引起的损伤

1. 紫外线引起的 DNA 损伤

紫外线(UV)照射引起的 DNA 损伤主要是形成嘧啶二聚体,DNA 分子最易于吸收的波长为 260nm 左右,故当受到大剂量的 UV 照射后,其一条链上的相邻的两个嘧啶核苷酸以共价键相结合,形成环丁烷嘧啶二聚体。相邻的两个 T 或两个 C 或 C 和 T 之间均可形成嘧啶二聚体,最容易形成的是 TT 二聚体,形成二聚体的反应是可逆的。较长波长的光(280 nm)有利于二聚体的形成,而较短波长的光(240 nm)有利于二聚体的分解。二聚体的生成位置和频率并非完全随机,而是与侧翼的碱基序列有关。在人皮肤中,每小时由于 UV 而产生嘧啶二聚体的频率为 5×10^4/细胞。UV 因穿透力有限对于人的作用仅限于皮肤,但可以影响微生物的存活。

2. 电离辐射引起的 DNA 损伤

电离辐射对 DNA 的损伤有直接效应和间接效应两种途径。前者指辐射对 DNA 直接沉积能量,并引起理化改变;后者指电离辐射在 DNA 周围环境的其他成分(主要是水)上沉积能量,而引起 DNA 分子的变化。水是活细胞的主要成分,水经辐射解离后可产生许多不稳定的具有高活性的自由基,进而引起 DNA 的损伤。

暴露于电离辐射后,DNA 的碱基和糖都可能发生一系列的化学变化,从而引起碱基的破坏和脱落,脱氧戊糖的分解。碱基的变化主要是由 OH^- 自由基引起的,生成各种过氧化物并发生咪唑环的破坏,一般说来嘧啶碱比嘌呤碱敏感。电离辐射照射细胞,由于细胞中 DNA 呈双螺结构,碱基损伤程度较轻,表明游离碱基又比核苷酸中的碱基敏感。OH 自由基还作用于脱氧戊糖,脱氧戊糖上的每个氢原子和羟基上的氢都能与 OH^- 反应,从糖基上夺去氢原子,脱氧戊糖分解,最后引起 DNA 链断裂。

DNA 受到电离辐射的另一个严重的生物学后果是链断裂。链断裂包括双链断裂(double-strand break, DSB)和单链断裂(single-strand break, SSB),双链中只有一条链断裂,称为单链断裂,若两条链在同一处或紧密相邻处同时断裂则为双链断裂。脱氧戊糖的破坏和磷酸二酯键的水解均会引起链断裂。另外,碱基的破坏或脱落也可间接引起链的断裂。

DNA 大分子受到照射后,SSB 约为 DSB 的 10~20 倍,在有氧条件下,DNA 的链断裂数增加,此外照射的剂量也将影响 DNA 的链断裂数。DNA 链断裂数随照射剂量的增大而增加。

电离辐射除了能引起DNA的碱基损伤和链断裂外，还能引起DNA交联，包括DNA链间交联和DNA-蛋白质的交联。DNA分子中一条链上的碱基与另一条链上的碱基以共价键结合，称为DNA链间交联(DDC)。DNA与蛋白质以共价键结合，称为DNA蛋白质交联(DPC)，目前对DNA 蛋白质交联研究得比较清楚，DPC的形成主要是由于OH^-的作用，氧含量、温度及染色质的状态对DPC的形成都有影响。体内DPC形成时对DNA和蛋白质还具有选择性，富含转录活性的DNA经电离辐射后易形成DPC。真核细胞中与DNA交联的蛋白质主要有组蛋白、非组蛋白、调节蛋白、拓扑异构酶及与复制转录有关的核基质蛋白等。5种组蛋白中形成DPC的反应能力也各不相同，组蛋白H3＞H4＞H2A＞H2B，而H1不与DNA交联。

三、化学因素引起的DNA损伤

1. 烷化剂对DNA的损伤

烷化剂是一类亲电子的化合物，极容易与生物体中的有机物大分子的亲核位点起反应。当烷化剂和DNA作用时，就可以将烷基加到核酸的碱基上去。DNA中的亲核位点主要有：腺嘌呤中的N_1，N_3，N_6，N_7；鸟嘌呤中的N_1，N_2，N_3，N_7，和O_6；胞嘧啶中的N_3，N_4和O_2，胸腺嘧啶中的N_3，O_2和O_4。其中鸟嘌呤的N_7位和腺嘌呤的N_3位最容易被烷化，DNA链上的磷酸二酯键中的氧也容易被烷化。烷化剂有两类，一类为单功能烷化剂，如甲基甲烷磺酸，只能与一个碱基起作用，形成单加合物；另一类为双功能烷化剂，如氮芥，能同时与DNA中两个不同的亲核位点反应，如果这两个位点在DNA双螺旋结构中的同一条链上，可产生一个链内交联，若两个受作用的碱基位于两条核苷酸链上，则形成了DNA的链间交联。鸟嘌呤对烷化剂中的磺酸乙基甲烷(ethyl methane sulfonate，EMS)等都特别敏感。嘌呤环的N_7位被烷化后，会在该部位形成四价氮。这个四价基团会促使嘌呤环离子化。离子化形式的烷基化鸟嘌呤可以与胸腺嘧啶配对，而不是与胞嘧啶配对，结果使G-C配对转换成A-T配对，造成DNA的损伤。烷化了的鸟嘌呤其糖苷键很不稳定，该键的裂解会导致碱基的脱落而形成DNA上的无碱基位点。此位点在复制过程中任何碱基都有可能插入，从而有可能造成碱基对的转换或颠换。DNA链上的磷酸二酯键被烷化则形成不稳定的磷酸三酯键，可能在糖与磷酸间发生水解作用，导致DNA链的断裂。

2. 碱基类似物对DNA的损伤

碱基类似物是一类结构与碱基相似的人工合成的化合物，由于它们的结构与碱基相似，当它们进入细胞后，便能替代正常的碱基而掺入到DNA链中，干扰了DNA的正常合成。最常见的碱基类似物是5-溴尿嘧啶(5-BU，见图13-1)，它在胸腺嘧啶环上有一个溴残基而不是甲基，与U的结构非常相似，能与A配对，也有酮式和烯醇式两种状态，当5-BU以烯醇式状态存在时，可与G配对，而且在烯醇式状态的频率更高，因此，一旦掺入到DNA链中，通过互变异构在DNA的复制过程中产生突变，引起A-T⇒G-C的转换。另一个常见的碱基类似物是2-氨基嘌呤(2-AP)，在正常的酮式状态时与T配对，在烯醇式状态时与C配对。

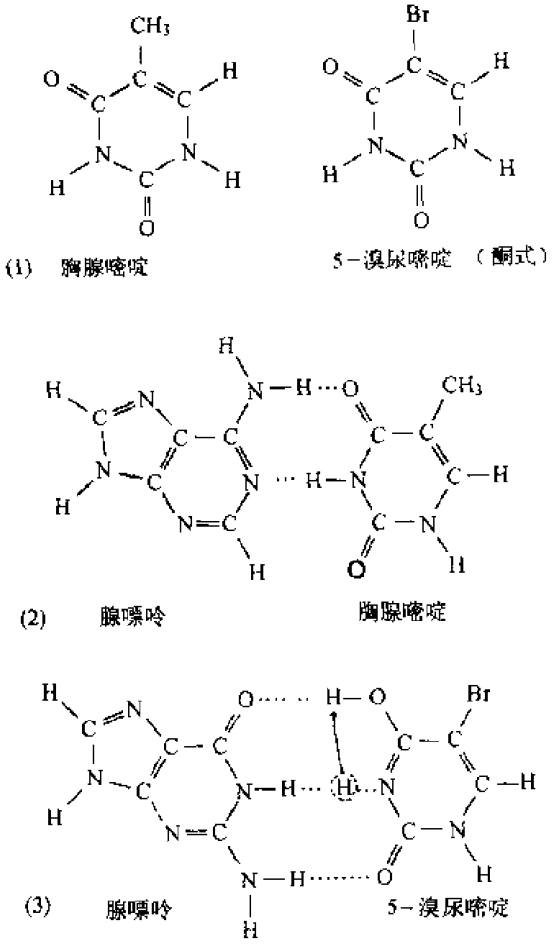


图 13-1 碱基类似物 5-BU
(1)T 和 5 BU 的结构相似；(2)A 和 T 的正常配对；
(3)A 和 5-BU 的配对。

另外，植物中能合成大量的毒性化学物质，其中有些物质是 DNA 损伤剂，这些 DNA 损伤剂有些还包含于人类的食物中，当摄入体内后将会对 DNA 造成损伤，但是其对人 DNA 所造成的损伤频率还不清楚。

第二节 DNA 修复[14]

显然，为了保证遗传信息的高度稳定性，生物细胞在进化过程中形成了一系列多步骤的修复机制。由于历史的原因，对 DNA 损伤和修复的研究仅是由那些对辐射——生物学反应感兴趣的生物学家进行。结果，这些课题与生物科学家们对复制、转录、基因调控和细胞周期等的兴趣相比很小。事实上，DNA 损伤和修复作为正常细胞生理过程的一部分，应当引起足够的注意。现在人们对细胞内的各种基本的过程如细胞周期、复制、转录、重组、修复等等有一个观念上的突破，就是认为这些过程都是相互作用、相互联系的。这种整体观念将有利于进一步理解细胞和有机体，并且对探索更好的癌症治疗途径和延缓衰老方法有积极意义。

DNA 修复研究最初可以追溯到 1935 年报导的紫外线照射大肠杆菌的实验，提出了损伤的“回复”过程(recover process)。随后 1949～1974 年间 Setlow 和 Carrier 等多人的研究揭示了所有的三大类细胞对 DNA 损伤的反应如损伤的回复、损伤切除、损伤耐受；1974～1984 年间主要揭示了 DNA 修复的机制；而 1984～1994 年期间主要进行了 DNA 修复研究。

一、切 除 修 复[22]

(1) *碱基切除修复*(base excise repair, BER)　一般地说，BER 可以去除因脱氨基或碱基丢失，无氧射线辐射或内源性物质引起的环氮类的甲基化等因素产生的 DNA 损伤。因此 BER 是 DNA 维持稳定的重要修复方式。BER 过程中的重要步骤是连接发生变化的碱基和脱氧核糖-磷酸盐链间的 N 糖苷键的水解，从而切除该碱基。碱基释放过程是由 DNA 糖苷酶催化的(见表 13－1)。

表 13－1　真核生物中碱基切除修复所用的 DNA 糖苷酶和 AP 内切核酸酶

酶	哺乳动物基因	人染色体基因定位	哺乳动物蛋白大小	酿酒酵母基因	酿酒酵母蛋白大小	DNA 中的作用位点
U-DNA 糖苷酶	*UNG1*	12	227aa 25.8kDa	*UNG1*	395aa	尿嘧啶
3－甲基腺嘌呤 DNA 糖苷酶	*AAG/MPG*	16	294aa 32kDa	*MAG*	296aa 34.3kDa	3 甲基嘌呤、7－甲基嘌呤、某些乙基化碱基、乙烯腺嘌呤、次黄嘌呤
嘧啶－DNA 糖苷酶	——	——	~30kDa	——	~40kDa	异常的嘧啶残基
FaPy-DNA 糖苷酶	*hMMH1*	——	——	*FPG1*	~40kDa	formamido 嘧啶类、7,8,二氢 8－氧鸟嘌呤
胸腺嘧啶错配-DNA 糖苷酶	——	12q24	410aa 46kDa	——	——	T-G、U-G 错配
AP－核酸内切酶	*HAP1*、*APE*(人)、*APEX*(小鼠)、*BAP1*(牛)	14 q11.2－12	35.5kDa	*APN1*	300aa 40.5kDa	无嘌呤和无嘧啶位点、哺乳动物中与 *E. coli* 的 *Xth* 同源的酶、酵母中与 *E. coli* 的 *Nfo* 同源的酶

每一种糖苷酶针对不同种类的 DNA 损伤，在切除碱基后，形成无嘌呤或无嘧啶位点(apurinic/apyrimidinic sites, AP 位点)，由 AP 核酸内切酶在此位点切开(见图 13－1)，然后由核酸外/内切酶除去 5′脱氧核糖磷酸盐残基。同时由 DNA 聚合酶修补缺口，在大肠杆菌中是 DNA 聚合酶 I，哺乳动物是由 DNA 聚合酶 b 进行修补。它们能以 DNA 断端的游离羟基末端为引物，以 4 种脱氧核苷三磷酸为原料，利用对侧的互补链为模板进行聚合反应，合成补丁，最后由 DNA 连接酶将相邻的两个核苷酸连接起来。在人细胞中还发现

了分子量约为 50kDa 的能从 DNA 双链中除去糖-磷酸残基的 DNA 脱氧核糖磷酸二酯酶(dRPase),主要是在当 DNA 聚合酶产生了单核苷酸补丁时起作用。

在 BER 过程中当修复合成的补丁很短时(very short-patch mode),主要是 DNA 聚合酶 b(polb)起作用,polb 在酶的小 N 端区域有 dRp 裂合酶活性,可能在 dRp 去除过程中与 dRpase 水解酶相互补充。小鼠胚胎实验证实此酶对哺乳动物细胞的正常生命力维持有重要作用,如参与 DNA 复制。当修复合成的补丁较长时,则可能由 DNA polsd 或 polse 进行,与短补丁时相比,这类修复需要 PCNA 参与,较长的修复补丁可能是小规模的转录反应再加上 5′-3′方向的链置换,因此产生了一样的结构。

(2)核苷酸切除修复(neucleotide excise repair,NER)　NER 除了能修复 UV 照射形成的嘧啶二聚体以外,还能消除补骨脂素衍生物和嘧啶的加合物,各种顺铂(抗癌药)——嘌呤加合物,以及由多环类致癌剂如乙酰氨基芴的加合物。体外实验中,NER 能修复如合成的 AP 位点、O^6 甲基鸟嘌呤和错配的碱基对。但 NER 对上述的加合物并非同样有效,修复的效率取决于加合物与正常碱基对的差别。例如 254 nm UV 照射引起的最主要损伤-环丁烷嘧啶二聚体从基因组中消除的速度比位于其次的损伤——光产物慢 5～10 倍。DNA 中光产物结构的扭曲程度比环丁烷嘧啶二聚体高,其中原因还不清楚。

NER 对存活不是必需的,有些可遗传着色性干皮综合征患者完全缺乏这种机制,其患皮肤癌的概率约高出正常人 1000 倍,表明人类 NER 的主要功能之一是修复 UV 照射引起的损伤。NER 系统在已研究过的真核生物中都很相似,说明其在进化过程中高度保守。哺乳动物和酵母中参与 NER 的各种基因列于表 13-2。

表 13-2　人类和哺乳动物中参与核苷酸切除修复的基因

人类基因	人染色体图谱位置	酿酒酵母同源基因	裂殖酵母同源基因	产物	
				分子量(kDa)	功能
XPA	9q34.1	*RAD14*		31	结合损伤的 DNA
XPC	3p25	*RAD4*		106	结合 ssDNA 与 XPC 结合
HHR23B	3p25	*RAD23*		43	同上
XPG(*ERCC5*)	13q33	*RAD2*	*rad13*	133	DNA 核酸内切酶,用于 3′端损伤
ERCC4(*XPF?*)	16p13.13	*RAD1*	*rad16*	126	DNA 核酸内切酶组分,用于 5′端损伤
ERCC1	19q13.2	*RAD10*	*swn10*	31	同上
XPB(*ERCC3*)	5q13	*SSL2*(*RAD25*)	*rhp3+*	89	3′-5′DNA 解旋酶
XPD(*ERCC2*)	11p14 3～15.1	*RAD3*	*rad15+*	87	5′-3′DNA 解旋酶
P44	17p13	*SSL1*		44	TFIIH 亚单位
P62	1p35～36 1	*TFB1*		62	TFIIH 亚单位
P52		*TFB2*		41	TFIIH 亚单位
P34		*TFB4*		34	"类 SSL1"TFIIH 亚单位
RPAP70	17p13	*RFA1*		68	与单链 DNA 结合
RPAP32	1p35～36.1	*RFA2*		29	同上
RPAP14	7p21～22	*RFA3*		13.6	同上
DDB, *UV-DDB*	11q12～13.2			127	与损伤 DNA 结合

在大肠杆菌中，NER 的关键酶由三个亚单位组成[21]，分别由 *uvrA*、*uvrB*、*uvrC* 等三个基因编码，组成的酶称 ABC 核酸内切酶，分子量为 246kDa。该酶能识别包括嘧啶二聚体在内的较大损伤所致的 DNA 螺旋扭曲，然后在这三种蛋白质作用下切割损伤的 DNA 链。

表 13－3　大肠杆菌中切除修复必需的 6 种蛋白质

蛋白质	相对分子量(kDa)	基序	活性	修复中的作用
a. 参与切除的核酸酶亚基				
1. UvrA	$(104)_2$	a)步查 ATP 酶 b) 锌指结构 c) 亮氨酸拉链 d)UvrA 超家族	a)ATP 酶 b)损伤特异性的 DNA 结合 c)UvrB 结合 d)TRCF 结合	a)损伤识别 b)分子伴侣 c)TRC
2. UvrB	78	a)螺旋酶基序 b)TRCF 同源序列	a)潜在的 ATPase b) 潜在的"螺旋酶" c) 损伤特异性的 ssDNA 结合 d) 与 UvrA 结合 e)与 UvrC 结合	a)损伤识别(后) b) 解开双螺旋 c) 3′切割
3. UvrC	69	a)与 UvrB 有限的同源性 b)与 ERCC1 有限的同源性(40 个氨基酸)	a)非特异性的与 DNA 结合 b)UvrB 结合	a)诱导 3′切割 b)产生 5′切割
b. 修复合成				
4. 螺旋酶Ⅱ(UvrD)	70	螺旋酶基序	真螺旋酶	释放 UvrC 和切除单核苷酸链
5. DNA PolI	103		DNA 合成	修复合成，替代 UvrB
6. 连接酶	75		连接酶	连接

NER 的关键特征是对损伤的 DNA 链两端进行切割，分别是在损伤 DNA 的两侧，修复补丁的大小在真核生物为 25～30nt，原核生物如大肠杆菌为 12nt。该长度与两次切割的位置有关。准确的切割位置是 DNA 上含有一个单修饰物，如环丁烷嘧啶二聚体和补骨脂素单加合物的切割是在损伤的 3′端上第 5 至 6 个磷酸二酯键上，另一切口在 5′端离损伤第 22～24 个磷酸二酯键上，产生 27～32nt 长的切除片段。准确的切割位置随加合物而异，但都在 3′端的第 3～9 个磷酸二酯键上和 5′端的第 16～25 个磷酸二酯键上。切割的工具是两类不同的核酸内切酶。哺乳动物细胞是 XPG 蛋白(3′)和含 ERCC1 的复合物(5′)。在酵母中为 RAD2 和 RAD1－RAD10 蛋白复合物。切割过程中涉及部分地打开围绕在 DNA 损伤部位的复合物，这样在双链 DNA 和开放区域之间的界线处即可由相应极性并且结构特异的核酸内切酶切开。该模式示意图见图 13－2，但准确的机制尚不清楚。

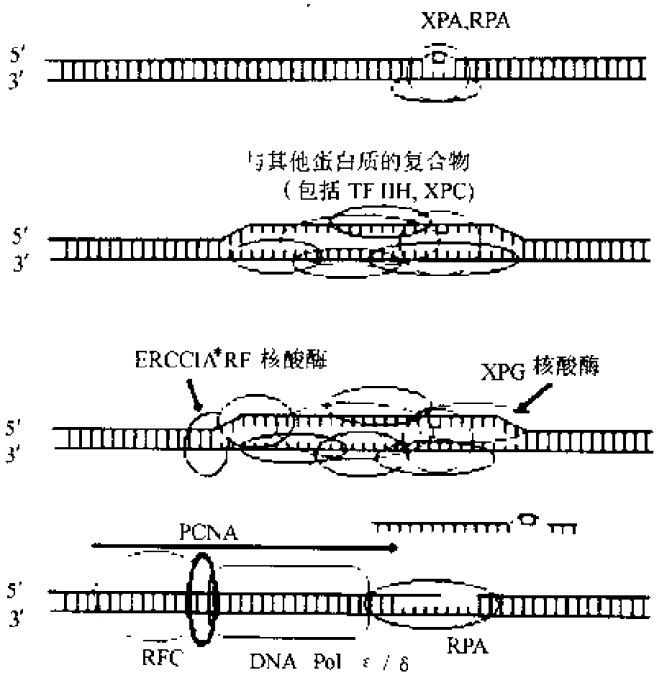


图 13-2　哺乳动物细胞核苷酸切除修复初步模式
从上至下依次为：损伤的识别、打开损伤附近的区域、两端切开、修复合成。

二、直接（回复）修复

为简单地把被损伤的碱基回复到原来状态的修复，可分为以下几种：

（1）O^6-甲基鸟嘌呤-DNA 甲基转移酶直接修复　O^6-甲基鸟嘌呤-DNA 甲基转移酶（MGMT）可以防止烷化剂所致的死亡和突变效应。存在于酵母和人类细胞中，通过从 O^6-甲基鸟嘌呤上把甲基直接转移到酶的半胱氨酸残基上来直接回复 DNA 损伤。此反应过程中修复酶失活。

（2）光解酶复活　酶学光复活过程是修复 UV 导致的环丁烷嘧啶二聚体的直接机制，这种修复有高度专一性。光解酶通过生色基团吸收蓝色或接近 UV 波长的射线，并且都含有还原 FAD 和第二个生色基团作为能量获取的“天线”，把能量转移给 FAD 辅助因子。被激活的 FADH 通过电子转移引发二聚体的裂解，把二聚体恢复到原先的单聚体形式，同时酶脱落。目前发现的第二类生色基团有二类：5,10-甲基四氢叶酸（MTMF）提供约 380 nm 处的最大吸收和 8-羟-5-脱氮核黄素（deazariboflavin），提供约 440 nm 处的最大吸收。

光解酶并非在所有生物中存在，如在裂殖酵母（*S. pombe*）、有胎盘哺乳动物（包括人类）的细胞提取物中都检测不到光解酶活性，有些植物中发现一些基因与微生物的光解酶有高同源性，但是没有光解酶或修复活性。植物的这方面修复还有待探讨。

另一种从果蝇胚胎中分离到的参与直接修复的酶能在光存在下从 DNA 上移去光产物，并恢复 DNA 活性，这种光产物是嘧啶的单聚体。此酶通过氧杂环丁烷中间产物而发

挥修复作用,其修复效率比环丁烷嘧啶光解酶低。在其他生物体内的生物学机能还不清楚。

(3)单链断裂修复　DNA单链断裂是损伤的一种常见形式,DNA单链断裂中有一部分是通过简单的重接而修复的,只需要一种酶——DNA连接酶参与,因此也属于直接回复。DNA连接酶能催化DNA双螺旋结构中双链之一的缺口处的5′磷酸根与相邻的一个3′羟基形成磷酸二酯键。

三、错配修复[23]

大肠杆菌中的错配修复系统　长补丁错配修复系统的典型例子为大肠杆菌的甲基化引导修复,这个系统已经在体外建立了精确的模型。该系统通过校正DNA生物合成过程中的错误,来稳定基因组。错配修复可以纠正几乎所有的错配,此外,对插入/删除引起的DNA链中多出1~3个多余核苷酸时也有修复作用。此类修复是以底物链上的信息为模板进行的,因此这个系统有区分底物链和新合成链的机制,细胞通过对底物链的甲基化来区分新合成的链。大肠杆菌中的错配修复系统至少包括9个蛋白成分,包括MutS、MutL与MutH蛋白质,它们分别在链的识别和修复过程中起作用。整个修复过程可以分为识别、切除和修补等步骤(如图13-3所示)。

识别的基础是甲基化,即Dam甲基化酶对(5′)GATC序列中的所有A的N^6位进行甲基化。由于发生在复制阶段后期,新合成的链上尚未被修饰,即新链上GATC序列中腺嘌呤A的未甲基化特征d(GATC)即可作为该链需进行校正的导向标记。仅需一个d(GATC)序列即可引导一个远在1 kb之外的校正过程,但当距离超过2 kb之后作用急剧减弱。先是MutS(约95kDa)与错配位点结合,然后MutL(约68kDa)在需ATP(非水解方式)条件下加到前者的复合物上。三者形成的产物激活与MutH蛋白

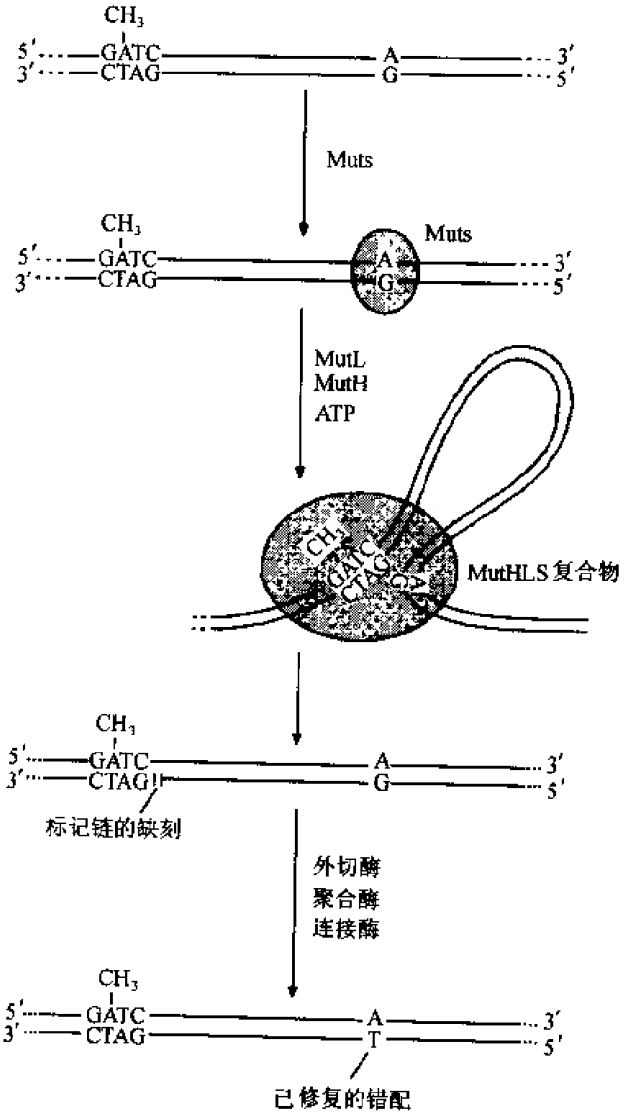


图13-3　大肠杆菌DNA错配修复的模型

(25kDa)结合的核酸内切酶，于 d(GATC)位点切开非甲基化链。链的裂口于是作为主要的标志引导对非甲基化链的修复。

切割由外切核酸酶水解进行，从断裂处开始并朝错配位点进行，终止于错配位点附近的几个不连续位点。由于 d(GATC)信号可以位于错配位点的两侧，反映了该修复系统的双向修复能力。但在反方向上修复时需要 MutS、MutL、MutU/vurD 和螺旋酶 II，并且需要不同的核酸外切酶。在错配的 5′方向上切除依赖 RecJ 外切酶或外切酶 VII，而在 3′方向则依赖核酸外切酶 I。由于这些外切酶都作用于单链，故推测螺旋酶 II 可能解开切割链利于外切酶切割。

最后一步是由 DNA 聚合酶 III 把空缺处补上，并连接修复产物。

四、交 联 修 复

链间交联是由多种多样的致癌剂和化疗药如亚硝酸、氨芥、补骨脂素等等所引起。在大肠杆菌中，链间交联修复(crosslink repair)按三个基本步骤进行：

1) 由(A)BC 核酸外切酶在一条链两端切开；

2) 由 RecA 介导同源性地与一个姐妹染色体配对；

3) 由(A)BC 核酸外切酶在两端切开另一条链。

哺乳动物体细胞内也有这种修复机制，甚至人的细胞提取物中就存在这种机制。但动物细胞内的修复过程细节还不清楚。

五、双链断裂修复

双链断裂(DSB)发生在体细胞重组和转座的生理条件下。此外，是电离辐射和氧化应激的主要损伤形式，如不及时修复则可能引起染色体丢失或重组。但是由重组产生的 DSB 和辐射所产生的不同：体细胞重组步骤之一有信号序列参与，故信号接点形成可认为是位点特异性的重组过程。编码接点的形成与正常的和酶学的 DSB 修复过程很相似，二者都有同源重组的酶参与。已知参与 DSB 修复的哺乳动物基因有：① *DNA-PK* 基因，② *XR-1* 基因，③ *XRCC1* 基因。DSB 修复的过程简单地描述就是：把末端接到一起，切齐，按需要产生 3′-OH 和 5′-P 末端，然后连接两个片段。

目前认识到依赖 DNA 的蛋白激酶(DNA-PK)对哺乳动物细胞中的双链断裂重接是必需的。DNA-PK 由一巨型的 465kDa 催化亚单位和 DNA 结合组分 Ku70 和 Ku80 组成，当结合到断裂 DNA 链上时被激活，并能使蛋白质底物磷酸化。在酿酒酵母中，双链断裂修复是由同源性重组完成的，有 RAD50－57 组基因产物参与。

六、转录和修复

从目前的研究中来看修复发生于转录的整个过程，如激活、起始和延长各个过程都与修复有关。在转录激活过程中，主要是 AP－1 和 NF-KB 两种转录因子可能参与转录-偶联修复。在转录起始过程中转录辅助因子 TF11H 也是一种修复因子，它由 7 个亚基组

成，其中的 XPB(p89)和 XPD(p80)亚基参与修复过程。转录延伸过程中转录的序列修复比非转录序列更迅速，在大肠杆菌中此阶段的修复过程是由 *mfd* 基因编码的转录-修复偶联因子 (TRCF)识别损伤部位的 RNA Pol 复合物并释放之，同时重新激活(A)BC 外切核酸酶的 UvrA 亚基。人类的修复基因也类似，但更复杂。

七、修饰和限制系统[6,13,15~16,18]

甲基化在真核生物有着不同的目的；即在不同功能条件下区分基因，而不存在与细菌相应的甲基化系统。菌株的宿主特异性是对 DNA 进行不同修饰的结果。反过来，这种模式可鉴别 DNA 的来源。修饰使细菌可以区分自体和"外来"(foreign)DNA，因为外来 DNA 无相同的修饰模式。这种区分使外来的 DNA 易被限制酶识别到其缺乏适当位点的甲基而切割掉。

这种限制酶修饰系统在细菌中普遍存在，其目的是保护自体 DNA 不被外源基因污染。一种细菌含有几套保护其 DNA 免遭外界物质或因复制错误引起的损伤。任何引起 DNA 结构产生与正常双螺旋不同的变化都会被纠正。变化可能是一个碱基变为另一种的单点突变，产生不按 Watson-Crick 规则配对的碱基对，或者是把一个庞大的片段加到 DNA 序列之中。

损伤部位被特殊的核酸内切酶识别并切除损伤，由其他酶合成替代序列，合起来称为修复系统，此外，还有处理复制损伤 DNA 后所产生的副作用的挽回系统。

这些系统包括一些能识别 DNA 中的结构或序列的酶。这些限制酶与 DNA 上的特殊靶序列结合，提供了另一种对蛋白质-核酸相互作用的认识。一些酶与 DNA 在某一位点结合，但在很远的另一位点切开 DNA，看来蛋白质可以在双螺旋 DNA 链上移动。一些修复酶明显是通过 DNA 双螺旋链上的非正常扭曲而识别 DNA 上的损伤位点。参与重组的酶可能结合两个分子的 DNA 提高它们之间的碱基配对能力。

以上我们可以认识到蛋白质和 DNA 相互作用的方法，这是基因分子生物学的关键之一。从这里我们可以得到对该系统的本质的总体面貌的了解。即保持和保护 DNA 以对抗环境或其他细胞系统的侵入。

修饰与限制系统是在对噬菌体感染效应基础上发现的。从一株细菌中释放的噬菌体 DNA 能成功地进行相同菌株中另一个细菌的感染，因为其与宿主 DNA 有相同的修饰模式。但从一个菌株上来的噬菌体 DNA 在转染到另一株时可被限制酶攻击。噬菌体好象被"限制"到仅一株菌上。

限制不是绝对的，一些感染性的噬菌体因为靶位点的变异或被宿主系统忽视而能逃避修饰。质粒对细菌菌株的特异性起作用。一些质粒(和噬菌体)具备限制与修复系统的基因，所以可以决定其特异性。由于这些系统存在，可对因质粒或噬菌体传递过来的细菌序列起筛选作用。

这种系统的基本特征是一个菌株内具有与限制活性相同的 DNA 甲基化酶活性。甲基化酶把甲基加到与限制性内切酶结合位点相同序列的 A 或 C 残基上。甲基化使得靶位点能抵抗限制酶切割。因此甲基化能保护 DNA 对抗切割。

限制性内切酶都可识别特异的、相当短的 DNA 序列作为结合位点。结合之后并在

识别位点处或其他部位切割，切割反应在体外测序和重构 DNA 中广泛应用。

由原有的细菌 DNA 在适当的位点甲基化后，可以对限制性活性发生免疫作用(图 13-4)。否则，限制性内切核酸酶将降解之。但这种保护作用并不对任何进入的 DNA 都有效。这种 DNA 具有未改变的靶位点，因此可被限制酶降解。限制与修饰的组合使得细胞可以区分外来 DNA 与其自身的受保护的 DNA 序列。这里"外来"DNA 指任何在菌株内的缺乏相同的修饰和限制活性的 DNA。在 DNA 的类型之间没有区别。同一菌株内的所有 DNA 都有相同的修饰模式。限制与修饰系统对 DNA 在同一菌株不同菌体之间的 DNA 传递不限制。

限制性酶可分为两大类，还可继续分类，主要的区别是第 II 型系统由 2 个酶承担修饰功能，而另一个承担限制功能。在第 I，II 型系统中仅一个酶兼备两种活性，其中 II 类酶最常见(表 13-4)。

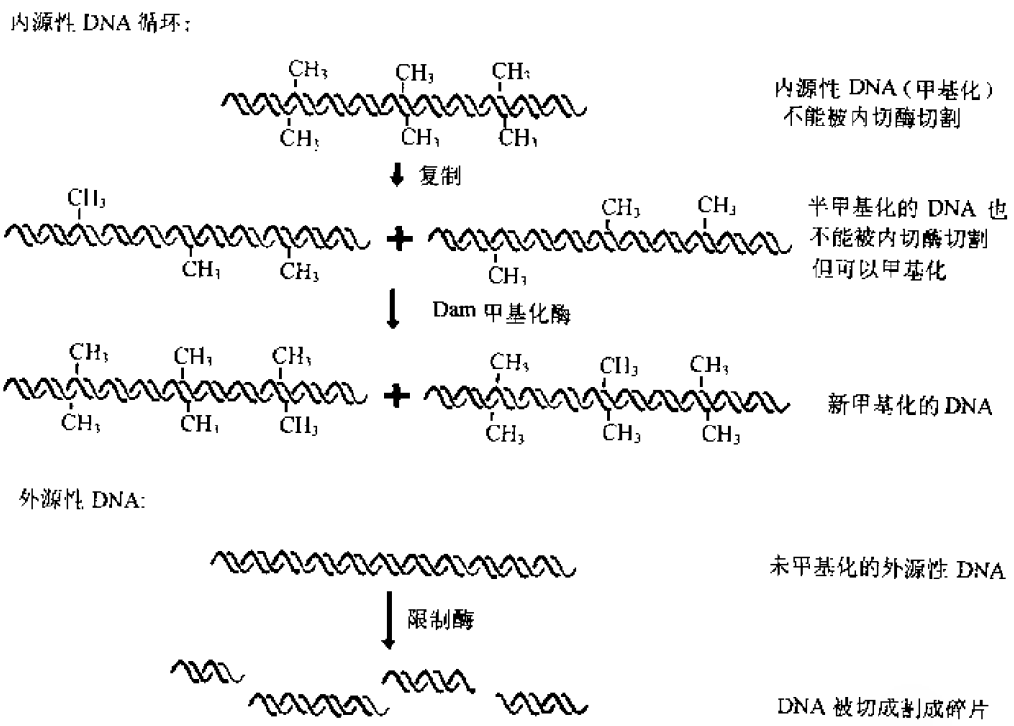


图 13-4 甲基化位点的防护作用

表 13-4 细菌中限制与修饰系统的酶

	II 类酶	III 类酶	I 类酶
酶蛋白质结构	内切酶和甲基化酶分离	一个酶有两个亚基和两种功能	一个酶两种功能，三个亚基组成
识别位点	4~6 bp 序列，通常为回文结构	5~7 bp 长的不对称序列	二分不对称序列
切割位点	与识别位点相同或接近	位于识别位点下游 24~26 bp	距识别位点 1000 bp 以上
限制或甲基化反应	分别反应	同时竞争	相互排斥
需 ATP	不	是	是

1. II 类限制酶

II 类限制酶很有用，常用于遗传工程。因为它们能识别很多特定的靶序列。每三种细菌菌株中就有 1 种以上 II 类系统。每一种酶仅负责限制作用。每一独立的酶负责同一序列的甲基化反应。

此类酶的结构不复杂。最特殊的例子是 *Eco*RI 系统成分，其限制酶由相同亚基构成一个二聚体，甲基化酶是一个单聚体(相同靶序列的识别可能是由复制识别 DNA 的蛋白质基序同向进化的结果)。

II 类酶的靶位点通常由 4～6 bp 构成。因为对称，两条链上的碱基都可以发生甲基化。这样位点可以被完全甲基化(两条链都改变)、半甲基化(仅一条链改变)或非甲基化。如图 13 - 4 所示。[9~10,20]

1) 完全甲基化位点 既不是修饰位点，也不是限制位点；

2) 半甲基化位点 不能被限制酶识别，但可被甲基化酶转变为完全甲基化；

3) 非甲基化位点 在体外可以是限制酶底物，也可以被修饰。细胞内，此类位点倾向于被限制；未修饰 DNA 被新宿主修饰而生存的可能性很小；

在细菌中，大多数甲基化过程只与当前的修饰状况的保持有关。完全甲基化 DNA 复制之后产生半甲基化 DNA。半甲基化位点识别可能是体外甲基化酶的通常作用模式。

大多数 II 型限制酶在非甲基化位点切割 DNA。每一条 DNA 链切开一个键。有些酶产生不整齐的切口，有些产生平齐末端。一种酶的亚类可以在识别位点错开几个碱基切开 DNA 链。

甲基化酶一次只加上一个甲基。如果底物均是非甲基化的，酶则仅引入一个甲基并从 DNA 链上脱落。

2. I 类酶[19]

第 II 类修饰和限制活性由 II 类和 III 类酶组成，都包含有多聚体和同时具有核酸内切酶活性和甲基化功能。二者作用机制与 II 类活性颇为不同。还不清楚 I 类或 III 类系统在细菌中到底占多大比例，但是肯定比 II 类系统少。

I 类酶以大肠杆菌 K、B 株的 hsd 系统中 *Eco*K 和 *Eco*B 变体为代表，并且最先被发现。此类系统由三种亚基组成(见图 13 - 5)，*Eco*K 酶由两个拷贝的 R 亚基，两个拷贝的 M 亚基和一个拷贝的 S 亚基组成。R 亚基负责限制功能，M 亚基负责甲基化，S 亚基负责识别 DNA 上的靶信号。当靶位点由 S 亚基识别后，酶结合到 DNA 上，随后进行限制或者修饰，因为 R 和 M 亚基在作用时是相互排斥的。

每一亚基都由一段基因编码。*hsdR* 突变体为表现型 r^-m^+，它们不能限制 DNA，但可以修饰之。*hsdS* 突变体为 r^-m^-，既不限制也不能修饰 DNA。*hsdM* 突变体可阻止限制或修饰作用。因此，M 亚基参与 R 亚基功能。这可能是一个失败——安全机制，使 hsdM突变体为 r^-m^-，而不是 r^+m^- 表现型，否则是致死性的(因为此时限制活性可降解细胞自身的非修饰 DNA)。

*Eco*B 和 *Eco*K 酶为等位基因的产物。他们的靶序列不同，但在二倍体中，同一菌株中的 S 亚基可能诱导另一菌株的 R 和 M 亚基活性。证明识别步骤与随后的切割或甲基

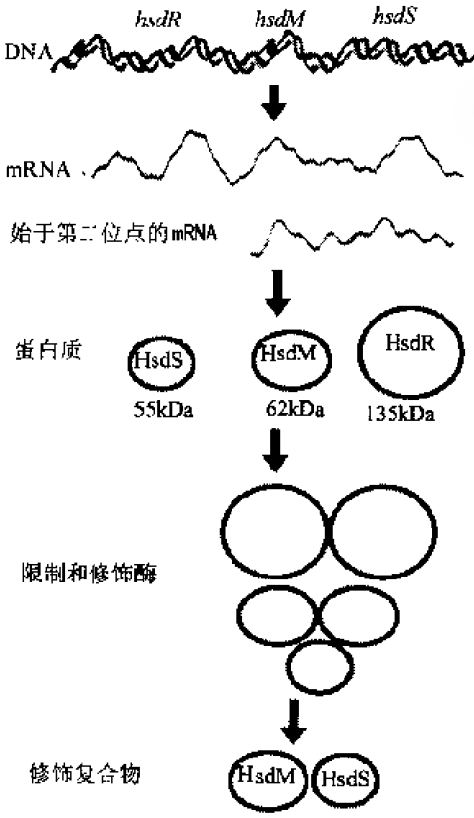


图 13-5 Ⅰ类酶的限制、修饰和识别亚单位

化过程无关。

*Eco*B 酶，具有相同大小的亚基，但它们的摩尔数比不同。B 株的 M 和 S 亚基可形成 1:1 复合物，可以与限制功能无关地行使甲基化功能。这可能是一个自然过程，因为三个基因位于一个操纵子内，顺序是：*P1-hsdR*-P2-*hsdM*-*hsdS*，其中 *P1* 和 *P2* 为独立的启动子。因此 *hsdM* 和 *hsdS* 可以独立于 *hsdR* 得到表达。

*Eco*B 和 *Eco*K 的识别位点为两部分结构，每一位点由一特异的 3 bp 序列和另一部分的 4 bp 序列组成，中间由几个碱基对分隔。识别位点分为两部分意味着两部分都位于 DNA 表面；中间插入区域的序列不重要。

靶位点的每一边看来被 S 亚基的特定区域所识别。HsdS 的基因之间的重组可以是两个区域的交互重组。重组 S 亚基利用从亲代菌株来的左边组分和另一亲代菌株来的右组分识别靶位点。识别位点的两侧都是不对称的，但二者都具有在互补链上修饰的腺嘌呤基团，如下列所示：

TGANNNNNNNNTGCT
ACTNNNNNNNNAGCA

某一位点的 DNA 是被切割或修饰是由靶位点的状态决定的，如图 13-5 所示。如果靶位点是完全甲基化的，则酶与之结合，但随即分离，不发生任何反应。如果靶位点是半甲基化的，酶则甲基化那条未甲基化链，以延续其在宿主 DNA 内的甲基化状态。如果靶位点为非甲基化，识别过程激活切割反应。由于甲基化的基团由辅助因子 S-腺嘌呤-甲硫氨酸(SAM)提供，在反应中转变为 S-腺嘌呤-高半胱氨酸(SAH)。SAM 与 M 亚基结合。在反应的初期，SAM 起着一个别构剂(allosteric effector)作用，改变 S 亚基构象使之与 DNA 结合。

与 DNA 结合后，下一步是与 ATP 反应。如果酶与一完全甲基化位点结合，ATP 到达后，使之从 DNA 上释放。在一个非甲基化位点上，ATP 把酶转变成一种易于进行切割的状态。这取决于 R 亚基。ATP 水解对切割 DNA 是必需的。在限制步骤发生之前 SAM 从酶上释放。

Ⅰ类酶表现出一种识别位点和切割位点之间的超常关系：切割位点距离识别位点 1000 bp 以上。切割不针对某一特定的序列，但也不是完全随机的，因为有一些 DNA 区域经常被切割(识别位点和切割位点的差异意味着识别位点不能用标志 DNA 损伤的方法定义；靶位点是由它们在甲基化修饰来定义的)。切割反应本身包括两步反应：首先，一条 DNA 链被切割；然后另一条链在附近切开，在切割位点两端区域发生核酸外切(exonu-

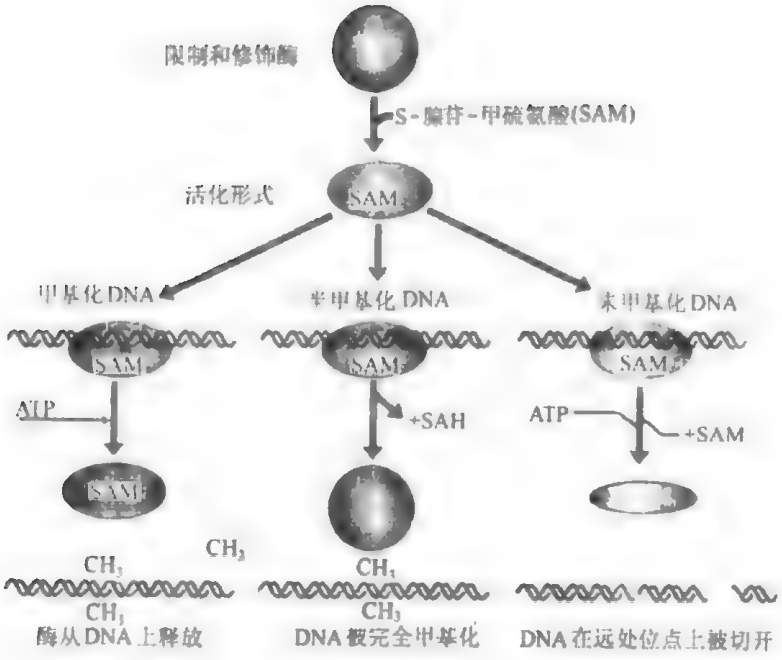


图 13-6　I类酶的作用机制

酶与靶位点结合，然后从完全甲基化的位点上释放出来，完成半甲基化位点的甲基化，或从非甲基化位点处沿 DNA 移动，切割该 DNA 分子的其他位点。

cleolytic)降解。ATP 大量水解，其功能还未知。

酶是如何识别一个位点，又在那么远处切割另一点呢？一个重要的特征是蛋白质从不放开其原先结合的 DNA 分子。如果酶与包括修饰和未修饰的 DNA 混合物一起孵育。在识别某一结合位点后，酶不与未甲基化 DNA 分离，以找到切割位点。

两种模型可以解释识别和切割位点之间的关系：①酶移动；②DNA 移动。见图 13-7。

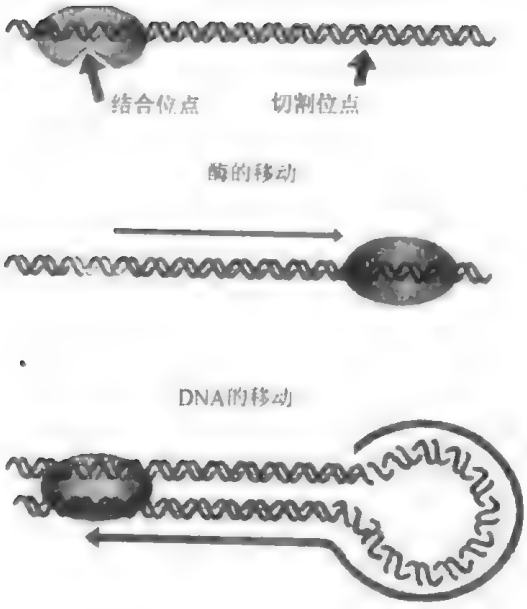


图 13-7　I类酶作用的两种模式

如果酶移动，它可能在DNA链上移动(原因未知)直到它实施切割。如果DNA移动，酶则可能仍然保留在识别位点上，由它扯动DNA通过酶上的第二个结合位点一直到达一个切割区域。图13-8显示了显微镜下酶产生的DNA环；看来酶切割之后仍然留在原来的识别位点上，支持第二种模型。

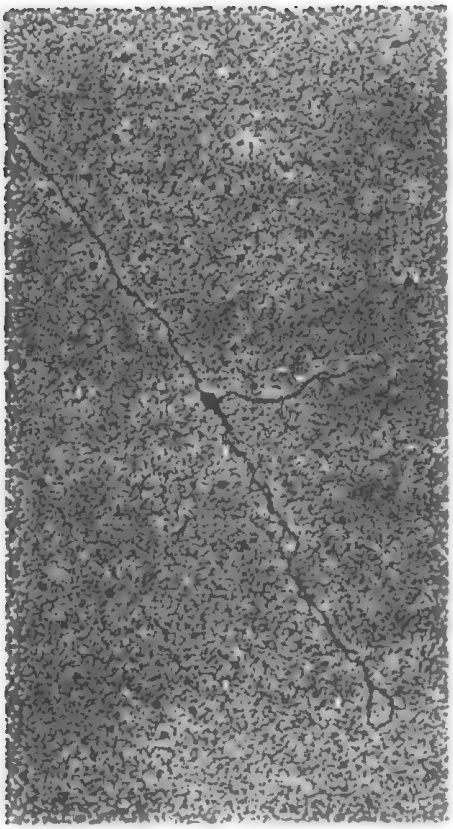


图13-8 I类限制酶作用实例

此酶可以同时结合DNA分子上的两个位点，产生一个环(Robert Yuan提供)。

3. III类酶[11~12]

有3种III类修饰和限制酶已经知道得较清楚[17]：*Eco*P1、*Eco*P15、*Hin*f。*Eco*P1和*Eco*P15是由大肠杆菌的质粒P1和P15编码，*Hin*f则见于血清型为Rf的人流感病毒，每种酶由两类亚基组成。R亚基负责限制功能，MS亚基负责修饰和识别。

修饰和限制活性同时表达。酶首先结合到其在DNA上的位点，结合时需要ATP(对ATP的依赖随酶种类而异)，然后甲基化和限制活性竞争与DNA反应。

甲基化发生在后位点，与MS亚基上有甲基化和识别机能的现象一致。限制性切割发生在一端的24~26个碱基远处，可能是酶足够大能达到这样的跨度，如图13-9所示。限制性切割产生不平齐切口，有2~4个碱基的差距。

酶甲基化腺嘌呤，但靶位点P1和P15具有一些内在特征：只能甲基化一条链。位点序列分别为：

sP1	AGACC	sP15	CAGCAG
	TCTGG		GTCGTC

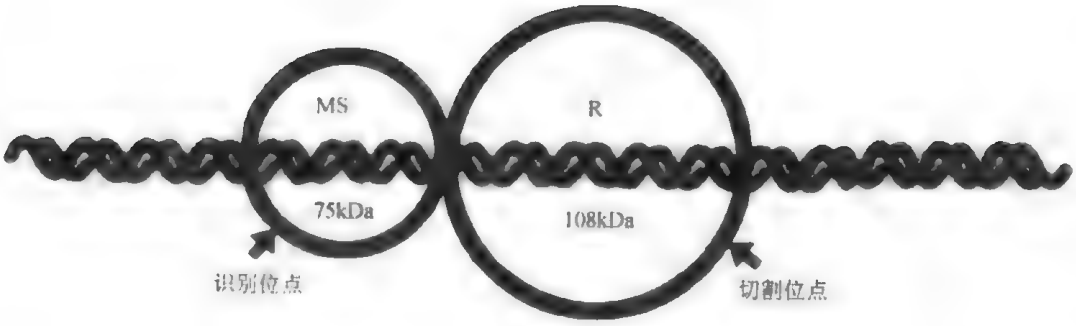


图 13-9 III类酶的两个亚基

(1)MS 亚基于靶位点上的识别和甲基化;(2)R 亚基于附近位点的限制性酶切作用。

甲基化状态是如何持续的呢? 图 13-10 显示了两类位点由复制产生。一个复制子具备原先的甲基化链;事实上它与亲代位点没有区别。另一复制子为完全非甲基化,所以原则上既是限制作用又是修饰作用的靶位点。是什么决定了它甲基化而不发生限制反应呢? 还不清楚,有一种说法认为修饰反应与复制作用相联系。

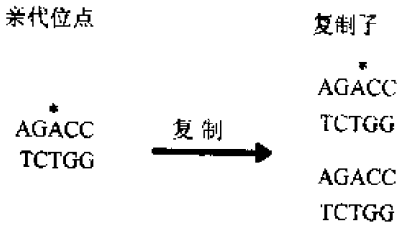


图 13-10 修饰与复制的联系

甲基化的(*)sP1 序列复制产生一个甲基化和一个非甲基化复制子。

单碱基改变 影响序列但不改变 DNA 总体结构。当 DNA 双螺旋分开时不影响转录和复制。这样这些改变将把损伤效应通过 DNA 序列上的改变传递给后代。

表 13-5 与 DNA 复制或修复有关的突变体基因

基因座	酶活性
mutD = *dnaQ*	DNA 聚合酶 III 的亚基
mutU = *uvrD*	DNA 螺旋酶(uvr 修复系统)
mutH, *U*	dam 修复系统的成分
mutL, *S*	Dtto; 也可以从 C-T 错配中移去 T
mutYa	从 A-G 和 A-C 错配中移去 A

八、大肠杆菌的挽回系统

挽回系统(retrieval system)亦称"后复制修复"(post-replication repair),因为它们在复制后起作用,或称为"重组修复"(recombination-repair),因为其修复过程与遗传重组过程相似。此系统在处理复制含有损伤碱基模板后产生的子代二倍体的缺陷中起作用。修复模型见图 13-11。

假如有一个 DNA 双螺旋结构扭曲,因为在双螺旋链上有嘧啶二聚体,当 DNA 复制

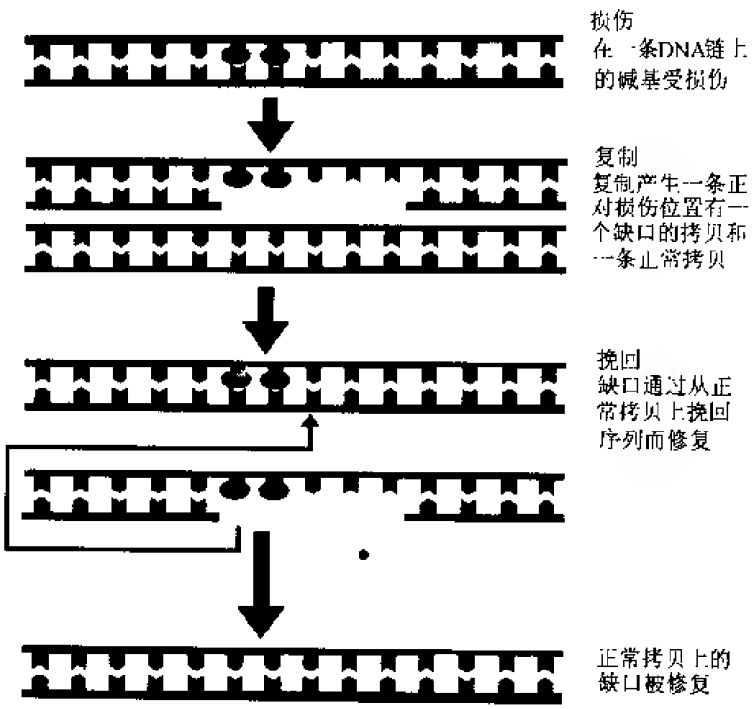


图 13-11　大肠杆菌的挽回修复系统

时，二聚体阻止损伤位点成为模板，复制只能被迫放弃这个位点。

DNA 聚合酶可能进行到达或接近嘧啶二聚体，然后聚合酶停止合成相应的子链，随后在更远的距离上重新开始复制，新合成的链上留下了一个空缺。

产生的子代双链在本质上不同。其中一条有父链的损伤加合物，另一条链上有一段长的空缺。另一对子链上则有未损伤的父链和一条正常的互补链。挽回系统正是利用了正常的子链。在含有损伤位点的那对子链中，与损伤位点相对应的空缺处是由获取正常子代双链中的同源性单链的信息而补上的。在进行单链交换(single-strand exchange)之后，受体双链就含有一条损伤的父链和另一条野生型链。供体双链有一条正常的父链而互补链有一个空缺；空缺可由通常途径的修复来合成填充，产生一条正常的双链。这样，损伤仅被局限于原先的扭曲位点。只要进行复制，就必须重复这个过程，直到损伤由切除-修复系统除去。

在缺乏切除修复机制的大肠杆菌中，*recA* 基因的突变基本上消除了所有的其他修复和修复机制。在 *uvr-recA* 细胞中试图复制 DNA 时产生与期望的 TT 二聚体之间距离相当大小的 DNA 片段。该结果提示当 RecA 功能缺乏时二聚体可产生一个致死性的复制障碍。这解释了为什么双突变体染色体组不能耐受 1～2 个以上的二聚体(野生型菌可以耐受达 50 个二聚体)。

recA 基因与其他的 *rec* 突变体一起鉴别出挽回途径。修复中所用的部分蛋白与重组中所用的相同，但不完全相同，因为有些突变体仅影响一种机能，而对另一种机能不影响。

recA 突变体几乎完全丧失遗传重组机能和挽回修复系统。RecA 蛋白具备在 DNA 分子之间进行链交换的能力，这是重组机能的关键，并且也参与重组-修复过程中的单链

交换(Single-strand exchange)。RecA看来也直接参与上述的倾向容错(error-prone)途径。

双突变体的特性提示 *recA* 参与两种 *rec* 通路。为检测是否有相似功能的两个基因是否为相同或不同通路的一个组分,把单点突变体的表现型与双突变体的表现型进行了比较。如果基因在同一通路中,则双点突变将与单点突变的表现型相同。如果在不同的通路中,双点突变将缺乏两种通路,而不是一种,这样将成为比单点突变更严重的表现型。通过这个设计判别出一个 *rec* 通路包括 *recBC* 基因,另一个包含 *recF* 基因。

recBC 基因编码核酸外切酶 V 的两个亚基,其作用由通路中的其他成分控制。在 *recA* 突变体中,酶不受控制,降解了大量的 DNA,即产生"卤莽的"表现型。*recF* 的功能还不知道,其他 *rec* 位点通过对重组-修复或重组的突变效应而得到鉴定。其中有几个是通过其对其他 *rec* 突变体的抑制能力而鉴别出来的。

这些基因的命名是基于突变体的表现型,但有时在一种条件下分离并命名为 *uvr* 基因座的突变体,结果却在另一条件下分离成为 *rec* 基因座。从这种不确定性得出了一个要点。不能定义有多少功能是属于一个通路或通路之间是怎样相互作用的,*uvr* 和 *rec* 通路并非完全独立,因为 *uvr* 突变体同时也显示重组-修复的效率降低。

可以预计今后会发现核酸内切酶、聚合酶和其他活性的网络,构建部分重叠的修复系统,或者在其中一种发挥某一功能的酶可被另一种其他通路的酶替代。

RecA蛋白直接参与重组-修复仅仅是其机能之一。这种特别的蛋白质还有另一种很特别的功能:当它被紫外线辐射激活(或另一种阻断复制的手段)时,可引起蛋白水解酶切割一系列的目标蛋白。原先认为 RecA 本身是一种蛋白酶,但现在看来是它触发了位于目标蛋白内潜在的蛋白水解活性。RecA 的激活负责诱导许多基因的表达,包括修复功能(如长补丁修复模式)。

RecA蛋白的双重活性使之很难知道 *recA* 突变体中的修复缺陷是由于丧失了 RecA 的 DNA 链——交换机能还是丧失了一些其他的依赖蛋白酶活性的机能的结果。

九、SOS 系统[7~8]

许多损伤 DNA 或抑制大肠杆菌复制的手段引起一系列综合的表现型改变称为 SOS 反应,这是由 RecA 蛋白和 LexA 抑制物相互作用而引起的。

由 UV 照射或交联剂或烷化剂引起的损伤以及多种方法导致的复制抑制如 T 丢失、使用药物,或基因发生突变都可以诱发 SOS 反应。

SOS 反应表现为修复损伤 DNA 的能力增强,这是通过诱导长补丁修复系统和 Rec 重组-修复系统组分的合成来实现的。此外,细胞分裂抑制,并可能诱导溶原性原噬菌体。

反应的初期是损伤处理激活 RecA,目前还不清楚损伤和 RecA 活性的突然变化有何关系。因为很多种损伤都可诱发 SOS 反应,目前的工作主要集中在 RecA 由一些普通的 DNA 代谢产物激活的想法上。

诱导信号可能由一个从 DNA 上释放的小分子组成;或是 DNA 本身形成的一些结构。体外 RecA 的激活需要单链 DNA 和 ATP 存在。这样激活信号可能是在损伤位点处的单链区域存在。不论信号是什么形式,其与 RecA 的作用很快:SOS 反应发生于 DNA 损伤后数分钟之内。

RecA 激活后可引起蛋白溶解性切割 LexA 产物。LexA 是一个小蛋白(22 kDa),在未经处理的细胞中相对稳定,对许多操纵子起阻遏物作用。切割反应有特别的地方:LexA 具有潜在的蛋白酶活性,并由 RecA 激活。当 RecA 激活,引起 LexA 发生自体催化的自身切割过程;这样使 LexA 阻遏功能失活,结果诱导了所有与之联结的操纵子。其过程见图 13-12。

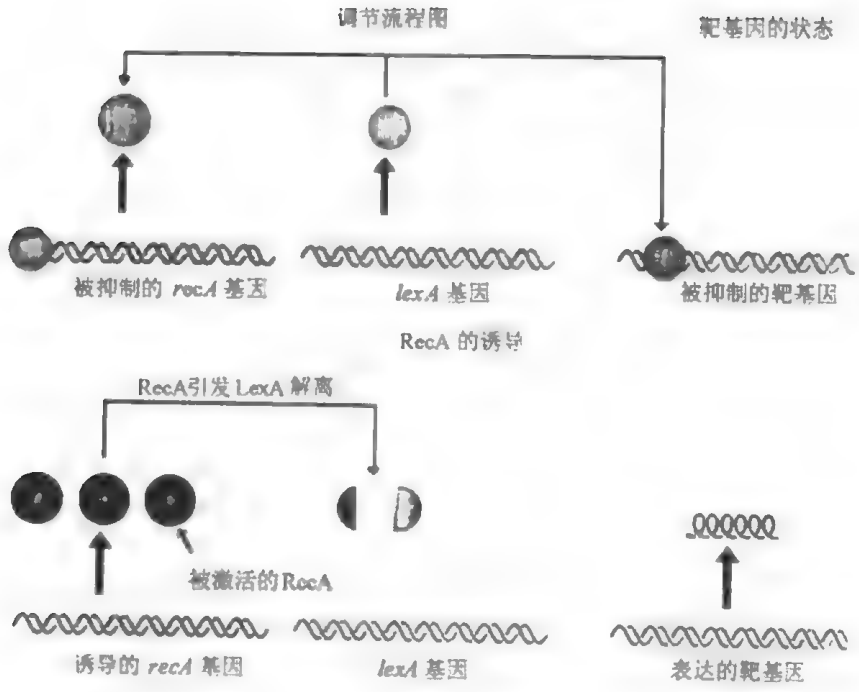


图 13-12 LexA 蛋白抑制许多基因,包括 recA 和 lexA 修复机能;激活 RecA 导致水解切除 LexA 从而诱导被抑制的基因。

LexA 阻遏的目的基因包括许多修复机能,目前其中仅一部分被鉴定出来,对 LexA 反应基因的系统扫描已经通过融合随机的操纵子和 *LacZ* 基因的方法构建,然后测定当细胞用 DNA 损伤剂处理时 β-半乳糖苷酶产物上升的水平。至少检出了 5 种相关基因,都称为 *din* (damage inducible)。原先已知的 SOS 反应组分的基因包括 *recA*、*lexA*、*uvrA*、*uvrB*、*umuC* 和 *himA*。

一些 SOS 基因仅仅在经处理的细胞中有活性,另一些则在未经处理的细胞中也有活性,但表达水平随 LexA 切除而升高。*uvrB*,为切除修复系统的基因组分之一,有两个启动子;一个与 LexA 无关,另一个受其调控。故 LexA 切除后,该基因可从第二个启动子处开始表达。

LexA 阻遏其目标基因是通过结合到一个 20 bp 长称为 SOS 盒(SOS box)的 DNA 序列上,该序列是对称的,并有一个拷贝存在于每一目的位点上。SOS 盒在不同的位点上并不相同,但在 8 个位点上完全保守。与其他操纵子相同,SOS 盒与各自的启动子重叠。在 LexA 位点,自行阻遏的主体有两个相邻的 SOS 盒。

RecA 和 LexA 在 SOS 环路中是两个相互关联的目标。RecA 激发 LexA 的切割,后

者阻遏 RecA。所以 SOS 反应同时放大 RecA 蛋白质和 LexA 阻遏物。结果不像开始时出现的那样矛盾。

RecA 蛋白表达增强是其在重组-修复途径中起直接作用所必需的。在诱导过程中，RecA 水平从其基线水平的约 1200 分子/细胞上升约 50 倍，诱导细胞中的高水平意味着有足够的 RecA 来保证所有的 LexA 蛋白被切开。以防止 LexA 重新形成对目的基因的阻遏。

但是该细胞环路的主要特性还在于其能迅速地回复到正常状态。当诱导信号被去除后，RecA 蛋白质失去使 LexA 脱稳定的能力。此时，*LexA* 基因高水平表达，在缺乏活化的 RecA 时，LexA 蛋白质迅速以未切割形式积累并关闭 SOS 基因。这可以解释为何 SOS 反应是完全可逆的。

RecA 也触发其他细胞靶点的切割，有时结果更直接。UmuD 蛋白在 RecA 激活后即被切割，以激活 UmuD 和倾向容错(error-prone)修复系统。

RecA 激活也引起其他一些阻遏蛋白的切割，包括几种原噬菌体的切割。其中有 λ 阻遏物(正是其中发现了蛋白酶活性)，这解释了为何 λ 总是由 UV 照射激活，蛋白溶解阻遏物被切除，释放噬菌体进入溶解循环。

这个反应不是一个细胞 SOS 反应，但代表通过原噬菌体识别细胞是否遇到麻烦。于是通过进入溶解循环来产生原噬菌体以保证存活。从这个角度来看，原噬菌体诱导反作用(piggybacking)于细胞系统是通过对同一诱导物(RecA)激活反应来实现的。

所有已知 RecA 激活时的靶蛋白都在 -Ala-Gly 二肽序列的中间多肽链上切开。在二肽的两端只有有限的氨基酸具有同源性，提示蛋白质的四级结构是目标识别的重要特征。

RecA 的两种活性是相对独立的，*recA441* 突变体使 SOS 反应发生而无须诱导，可能因为 RecA 自发地保持在激活状态。其他突变丧失了被激活的能力。两种类型的突变都不影响RecA处理 DNA 的能力。反向类型的突变使重组功能失活，但仍完全保留诱导 SOS 反应的能力，这样产生对 RecA 在修复通路中的直接或间接作用时的去干扰。

十、与修复有关的人类疾病

有几种人类的遗传疾病是由于 DNA 修复基因的突变所引起的：

(1) 着色性干皮病 (xeroderma pigmentosum) 此疾病是由于切除修复或者后复制修复的缺陷所引起。特征是光照性皮肤病和早期皮肤癌，有些病例表现神经系统异常。

(2) 科克因综合征 (Codkayne's syndrome) 特征为光敏感性、侏儒症和精神发育迟缓，典型的形式是由于 *CSA* 和 *CSB* 基因突变所致，这两种基因为转录-修复偶联(TRC)所必需，而其他形式则在 XPB; XPD 或 XPG 等的基因上发生突变。由于 XPB 和 XPD 是 TFIIH 的亚基，XPG 与 TFIIH 紧密结合，这些基因突变导致缺乏 TRC。引起该综合征。

(3) 运动失调性毛细血管扩张症(atacsia telangiectasia, AT) 此症是以小脑性运动失调，癌症发生率高和电离辐射(非 UV)敏感为特征的自发退化性疾病。原因是编码类 PI-3 激酶的基因 ATM 在所有的 4 个互补基因中发生变异。

(4) Bloom 综合征 (BS) 特征是中度的光过敏和侏儒症，精神发育迟缓。BS 细胞

的姐妹染色质交换速率升高，并且单链缺口处接合速度缓慢。

(5) 范可尼贫血(Fanconi's anemia, FA)　表现为再生障碍性贫血，各类血细胞减少，最终成为白血病和早亡。FA 细胞表现为染色体畸型并且对交联剂极其敏感。与此有关的有 4 个互补基因 *FA-A* 到 *FA-D*。其中 *FAC* 基因已被克隆，编码 63 kDa 蛋白质。

(6) 肌萎缩性(脊髓)侧索硬化症 (amyotrophic lateral sclerosis)　此症的特征是整个神经系统运动神经元进行性的退化。有些家族性的形式与 Cu/Zn-SOD 的错义突变有关。人们推测 SOD 减少使得过氧化物蓄积损伤运动神经元。

第三节　基 因 突 变[24~25]

在所有生物的基因组中的 DNA 并不稳定，相反，很容易发生各种各样的可以遗传的改变，即基因突变(mutation)。携带突变的生物个体或群体(株/系)称突变体(mutant)。较大规模的突变和染色体异常包括染色体的丢失或获得以及染色单体的断裂或重新接合，较小规模的突变可按不同的角度分为许多类，如仅一个 DNA 碱基发生突变，称点突变(simple mutation)，也有两个等位基因或非等位基因序列之间的交换。其中较重要的分类是按突变的形式，包括：① 碱基替换(base substitution)，② 缺失(deletion)；③ 插入(insertion)。碱基替换通常仅发生在一个碱基上，但偶尔也有几簇碱基同时被替换；缺失指一个或多个核苷酸从一个序列中被删除；插入指一个或多个核苷酸被插入某一序列，但也包括从另一位点的移换。拷贝或复制移换指一个位点上的序列被复制并被插入另一位点。非拷贝移换指 DNA 序列从一个位点简单地移换到另一位点。后者在人和哺乳动物染色体组中少见，而多数通过 RNA 介导故为另一位点的拷贝。

一、突变的命名和表示方法

人类基因和等位基因一般以人类基因图谱会议命名委员会所定的名称命名，等位基因的变异体收录在 OMIM 数据库中并且按顺序标号，方法是基因名称后跟着星号 (*) 和系列号，例如 HBB*0403。1995 年和 1998 年的 Trends in Genetics 副刊出版了包括许多生物的遗传学命名法。表 13-6 则是描述突变的命名法[26]。

表 13-6　描述突变的命名法

氨基酸替换
用单字母编码：A——丙氨酸；C——胱氨酸；D——天冬氨酸；E——谷氨酸；F——苯丙氨酸；G——甘氨酸；H——组氨酸；I——异亮氨酸；K——赖氨酸；L——亮氨酸；M——甲硫氨酸；N——天冬酰胺；P——脯氨酸；Q——谷氨酰胺；R——精氨酸；S——丝氨酸；T——苏氨酸；V——缬氨酸；W——色氨酸；Y——酪氨酸；X——终止密码子。
R117H——第 117 氨基酸从精氨酸变成组氨酸；
G542H——第 542 位甘氨酸由终止子替换。

续表

核苷酸替换

核苷酸标记通常在前面加上 nt（核苷酸）或 np(核苷酸对)。一般情况下号码为 cDNA 的有意义链(线粒体突变例外)。因为许多 cDNA 的 5′端未加定义，所以其编号有可能是已发表的特殊序列。内含子中的核苷酸可通过它们与最近外显子的相对位置来区别。

1162(G→A) 或 np1162(G→A) ——cDNA 中的第 1162 号鸟嘌呤替代；

621 + 1(G→T)——内含子中的第一个核苷酸(鸟嘌呤)，即位于一个外元的最后一个核苷酸(第 621 位)紧后边，被胸腺嘧啶替代。

1781-5(C→A) ——在 1781 位置 5′端前边内含子中第 5 个碱基——胞嘧啶替换为腺嘌呤。

缺失和插入

Delta-F508 或 ΔF508——第 508 号苯丙氨酸密码子缺失；

nt6232(del5) ——5 个核苷酸缺失，其中第一个位于 6232 处；

nt409(insC) ——在 409 位核苷酸后边插入胞嘧啶。

野生型(wild type)和突变型(mutant)是一个相对的概念，野生型是指没有发生突变的基因；而突变体就是指相对的某一性状的野生型基因发生突变之后，所表现出来的性状发生改变的表现型，其本质为 DNA 分子结构上的改变。在大多数情况下野生型表现为具备某种能力的性状(用 + 表示)，而相对的突变体表现为缺乏这种能力的性状(不用 - 表示)。如负责细菌合成精氨酸(Arg)能力的基因 *arg*，野生型表现型为 Arg^+，其突变型由于不具备合成精氨酸的能力，即为 arg^-。野生型和突变体的表示方法见表 13 - 7。

表 13 - 7　细菌突变体的突变表示法

表现型	
合成某种物质的能力有、无	Sub^+, Sub^-
抗生素耐受或敏感	Ant^r; Ant^s
基因型	
能合成物质的野生型	*sub*$^+$
上述基因型的变异体	*sub*$^-$
突变的 *subA* 基因；第 63 号突变(常用)	*subA*$^-$；*subA*^{63*}；*subA*$^{(TS)*}$；*subA*$^{(Am/Oc/Opal)*}$
对某种抗生素抗性或	
敏感的基因型	*ant*
对某种抗生素抗性；敏感性基因型	*ant*^{r}; *ant*^{s}

* 数字表示分离的时间顺序；(TS)表示温度敏感型突变；(Am)表示琥珀型无义突变(UAG)，(Oc)表示赭石型无义突变(UAA)，(Opal)表示乳石型无义突变(UGA)。

二、突 变 分 类

突变发生的过程称为突变生成作用(mutagenesis)。在自然界中发生的这个过程称为自发突变生成(spontaneous mutagenesis)，所产生的突变体称为自发突变体(spontaneous mutant)；人为地使突变发生的过程，称为诱变(induced mutagenesis)，所产生的突变体称诱发突变体(induced mutant)。诱变和 DNA 序列测定方法一样，都是分子生物学研究中的重要方法。

从不同的角度来看，基因突变可以分为许多相互联系的类别。从突变中碱基对变化

的数目上来看，可分为单点突变（point mutation）和多点突变（multiple mutation），单点突变指仅发生一个碱基对的突变(多为替代)。多点突变是指两个以上碱基对的突变。单点突变又包括碱基替代(base substitution)、碱基插入（base insertion）和碱基缺失(base deletion)（见图 13－13)。各种形式的突变的发生频率各不相同，人染色体组中各类突变发生率如表。

表 13－8　人类染色体组中各类突变的发生频率

突变类别	突变形式	发生率
碱基替换	所有形式	编码 DNA 中较普遍，且非编码 DNA 中常见
	转换和颠换	较少见，转换比颠换常见；尤其在线粒体 DNA 中
	同义和非同义替换	在编码 DNA 中同义比非同义替换常见；保守型替换比非保守型替换常见
	类基因转座过程（多个碱基替换）	罕见，但在某些串联重复位点或簇重复序列例外
插入	一个或数个核苷酸	在非编码 DNA 中常见，但在可能产生移码的编码 DNA 中罕见
	三个一组插入	罕见，但有些疾病由此引起，尤其是神经系统疾病
	大片段插入	罕见；偶尔可见大规模串联拷贝和插入转座成分
缺失	一个或数个核苷酸	在非编码 DNA 中常见，但在可能产生移码的编码 DNA 中罕见
	大片段缺失	罕见，但经常发生于含有串联重复序列或散布的重复序列之间
染色体异常	数量上和结构上	作为结构突变少见，但病理状态下常见，在体细胞突变更常见，并常见于肿瘤细胞

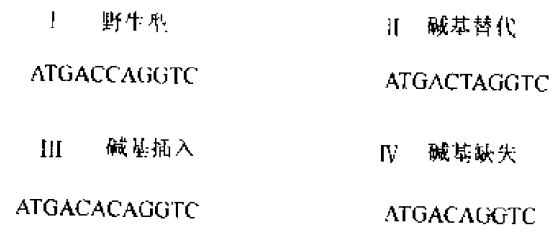


图 13－13　突变的三种替代（仅显示 DNA 单链的变化）

碱基替代是突变发生的常见形式，可以分为两类：① 转换（transition），即一种嘧啶(C/T) 替换另一种嘧啶或者一种嘌呤(A/G)替换另一种嘌呤。② 颠换(transversion)，即一种嘧啶被一种嘌呤替换或一种嘌呤被一种嘧啶所替换(见图 11－2)。从理论上来说，转换发生频率要比颠换发生频率高一倍，从一些研究数据上来看在一些产生氨基酸序列改变的突变中，转换与颠换的频率比略高于 2:1。

从突变发生的效应来看，有 ① 同义突变(synonymous mutation)（无声突变）不改变基因产物的序列；② 非同义突变即导致多肽产物的序列改变或功能性 RNA 的序列改变，序列中一个或多个成分被改变或去除，或者一般多余的序列插入产物之中。无声突变又可称为中性突变，因为对发生突变的染色体组既无益处，也无害处。而非同义突变可分为三类：一类对产物的功能无影响；第二类是带来好处的(如改善基因功能或基因-基因相互作用)；第三类是带来有害效应的。其中大多数非同义突变是有害的或致命的。然而由

于自然选择的结果，在群体中这类突变发生的频率减低。这样，在编码 DNA 中的突变率要远远低于非编码 DNA，于是，某一特异基因的编码部分和产生的氨基酸序列显示一相对高水平的保守性，同样一些重要的调节序列如启动子和增强子以及紧靠外边两侧的序列都是相对保守的。自然选择的压力使存在于编码 DNA 中的突变数目下降，也使发生在其中的突变种类减少。如缺失、插入一个或数个核苷酸的情况在非编码 DNA 中常见，但在编码 DNA 中几乎没有。因为这类突变导致翻译的可读框移动即移码突变(frameshift mutation)并且具有潜在的致病性。移码的结果是在突变下游的氨基酸序列与原先完全不同，而且多肽序列往往中间断开，因为提前碰上了终止密码子。

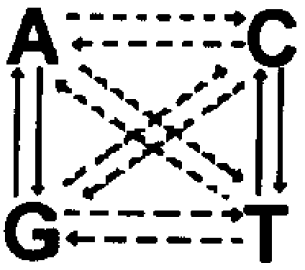


图 13-14 转换和颠换示意图
- - 颠换；—— 转换

无义突变是指非同义的替代使代表一个氨基酸的密码子被一个终止子替换，这种突变几乎总是导致基因功能的急剧改变，所以选择的压力迫使这类突变很罕见。错义突变是非同义替代导致代表另一个氨基酸的密码子取代了代表原先氨基酸的密码子。可分为两类：

(1)保守性替代　指突变导致一种氨基酸被另一种化学性质相似的氨基酸所取代，这样对蛋白质功能的影响很小。为减少核苷酸替代的影响，遗传密码看来是不断进化的，这样表示相关的氨基酸的密码很接近，例如 Asp(GAC,GAT)和 Glu(GAA,GAG)密码子对使得在第三个碱基的变动影响最小。但有时第一个碱基也有变动的余地，如 CUX (Leu)→GUX (Val)。

(2)非保守性替代 (nonconservative substitution)　指突变导致一种氨基酸被另一种具有不同侧链的氨基酸所替代。有时是所带电荷的差别，有时是极性与非极性侧链的区别。这种替代通常发生在密码子的第一和第二个碱基上，如 CGX(Arg) ? GGX(Gly)，CCX (Pro) ? CUX(Leu)等等。

从突变体对外界环境的敏感性来看，可分为条件型突变 (conditional mutation) 和非条件型突变(nonconditional mutation)。条件突变体表现为条件致死，最常见的是温度敏感型突变体 (temperature-sensitive mutant)。这种突变体在必需基因的研究中成为很重要的方法。温度敏感突变一般为错义突变。细菌温度敏感突变体的许可温度为 30℃ 左右，非许可温度为 42℃，少数突变体相反。

从突变体又恢复原先野生型表现型的突变过程称为回复突变(reverse/back mutation)。这个突变可能是其抑制了原先突变体的表现而仅使其表现型与最初的野生型表现相同，也可能是正好回复成为与野生型的基因型也相同，但显然后一种可能性较小。前者一般地说两次突变分别发生在两个位点上，其中第二次突变因抑制了从野生型到突变体的效应称为抑制突变(suppressor mutation)，如果此突变发生在正向突变的基因之内部，称基因内回复(intragenic reversion)；如果发生在其他基因之间，则称为基因间回复(intergenic reversion)。

突变热点 (hot spots of mutation) DNA 分子上每一个部分发生突变的频率并不相同，在某些位点发生突变的频率远远高于平均数，这些位点称为突变热点。突变热点的形成与造成突变的机制有关。

三、基因突变的后果

（一）功 能 丧 失

抛开生化学研究仅从遗传学的角度来看，可以看出某一表现型是由功能丧失或获得所引起的。因功能丧失导致的表现型，多数是由于某一基因突变后使其所表达的蛋白质失活。这种突变可以是点突变，也可以是缺失或基因紊乱所致。例如人的Ⅰ型 Waardenburg 综合征（MIM193500），是 *PAX3* 基因中的突变引起的氨基酸替换、移框突变、剪接突变或整个 *PAX3* 基因丢失引起。如图 13－15 所示。

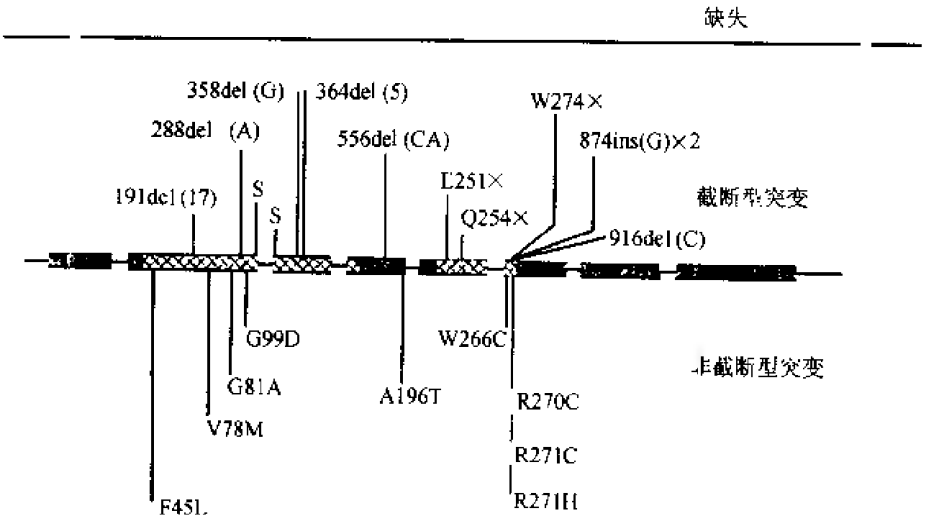


图 13－15 *PAX3* 基因的功能丧失型突变

PAX3 基因有 8 个外显子。与基因产物功能有关的关键序列用＝表示。其中 S 表示剪接位点突变。上半部分所示的突变将导致一个半截的基因产物，因为移框或剪接突变之后使翻译提前终止；下半部分所示的突变将产生一个全长的蛋白质，但有些氨基酸序列被替换；有些患者整个基因被切除。这些突变的表现型都相同即功能丧失。

功能丧失可能是部分或全部的。不编码产物的等位基因通常称为无效等位基因（null alleles）。有时基因产物没有活性，但能够用免疫学方法检测出来，表示蛋白质仍然能被合成和输出。这种突变常称为 CRM^+（positive for cross-reacting material）。

导致一蛋白质缺失的突变不一定位于编码这个蛋白质的基因之中。丙种球蛋白缺乏症是一种免疫缺乏症，一般都会推测是免疫球蛋白的基因发生突变引起的。但免疫球蛋白基因座位于 2 号、14 号和 22 号染色体上，而缺乏症突变不发生在这些基因座上。有许多是伴性遗传的。可能与一蛋白质合成过程中的加工和修饰有关，还有 B 细胞成熟等免疫系统的基本过程有关的异常都可导致免疫缺乏症。

一个基因缺失有时会引起多种酶缺乏。黏脂病Ⅱ型（MIM2525000）是由于多种的溶酶体酶缺乏。主要的缺失不是在这些酶的结构基因当中，而是对这些糖化的酶分子上的

甘露糖残基进行磷酰化的 *N*-乙酰葡萄糖胺-1-磷酸转移酶缺失，结果为一系列的溶酶体酶缺乏。

引起某一基因产物功能减弱或丧失的途径有许多种，列于表 13-9。

表 13-9 引起某一基因产物功能发生变化的因素

改变	例子
整个基因丢失	大多数的 α-地中海贫血；
部分基因丢失	60%的面肩肱型肌营养不良
基因结构紊乱	
(1)通过易位	肌营养不良妇女的 X-常染色体易位
(2)通过倒位	*F8* 基因倒位(血友病)
基因中插入一段序列	血友病 A 型的 *F8* 基因内插入 *LINE-1* 重复序列
抑制或防转录	*Fragile-X* 完全突变
启动子突变减低 mRNA 水平	β-球蛋白-29CA→G 突变
减弱 mRNA 稳定性	Hb Constant Spring (HB20001)
使供体的剪接位点失活 (引起通读到内含子)	*PAX3* 451+1(G→T)突变
使供体或受体剪接位点失活 (引起外显子被忽略)	*PAX3* 突变
激活潜在的剪接位点	β-球蛋白内显子 1-110(G→A)突变引起地中海 β-型贫血
引起翻译中的移框突变	*PAX3* 847 ins(G)突变
把一密码子转变成一终止密码子	*PAX3* Q 254X 突变
替换一个重要的氨基酸	*PAX3* R271C 突变
影响转录后的处理	家族性高胰岛素原血症(切割抑制)
影响产物在细胞内正确定位	ΔF508 突变引起的囊纤维变性

(二)功 能 获 得

有很多机制可以产生功能获得的显性表现型(见表 13-10)。但获得完全是新功能的情况很罕见，除了在癌症中；而在遗传性疾病中，功能获得往往是基因在发育过程中的错误时期、错误组织，或对错误信号的反应，或者是异常高水平表达的结果。

表 13-10 功能获得突变的机制

机能不良	基因	疾病	MIM 编号
获得新底物	*PI*	α_1-抗胰蛋白酶缺乏	107400
过度表达	*PMP22*	进行性神经性腓骨肌萎缩	118200
受体持续"接通"	*GNAS1*	McCune-Albright	174800
离子通道不适当"开放"	*SCN4A*	先天性肌强直病	168300
结构异常的多聚体	*COL2A1*	成骨不全	多种
嵌合基因	*BCR-ABL*	慢性骨髓白血病	15140
未知	*HD*	亨廷顿舞蹈病	143100

图 13-16 是一例罕见的先天性点突变导致一种蛋白出现全新的功能，此突变位于

Pittsburg 染色体的 *PI* 基因座上。而在癌细胞中由于外显子的打乱机制常见有新功能的基因，如许多肿瘤特异的染色体重排产生的嵌合基因 可以导致无法控制的细胞增殖。

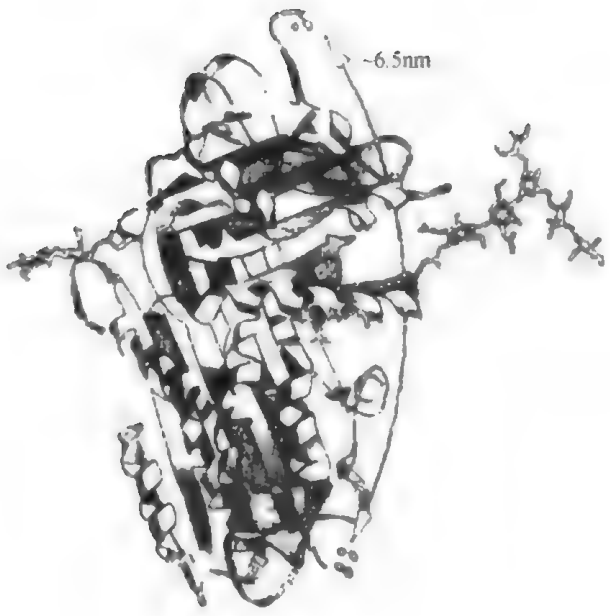


图 13－16　功能获得性突变的例子

α_1-抗胰蛋白酶分子能抑制弹性蛋白酶。第 358 位甲硫氨酸(活性中心)是弹性蛋白酶的"诱饵"；当 Met358 和 Ser359 之间的肽键被切开时，两个残基弹成相距 6 5nm 远。弹性蛋白酶失活。而 Pittsburgh 突变体有一个 M358R 的错义突变，蛋氨酸变成精氨酸。丧失了与弹性蛋白酶的亲和力，但是形成了对凝血酶的诱饵。新形成了抗凝血酶活性，此突变体是一种致命的出血性疾病(本图引自日内瓦大学 ExPASY 分子生物学互联网服务器)。

(三) 癌 症 发 生

体细胞突变频繁且不可避免，典型的人体细胞突变每个基因每代发生率为 10^{-5}到 10^{-7}，但是并非所有的突变都会导致疾病如癌症或遗传病。一个正常细胞转变成恶性肿瘤需要以多种特异性的突变同时发生为条件。最近的研究表明有三类基因突变常见于癌症，即癌基因(oncogene)、肿瘤抑制基因(tumor suppressor gene，*TS*)和增变基因(mutator gene)。

1. 癌基因

癌基因(oncogene)可以增强细胞的增殖。在正常的非突变体基因为原癌基因(proto-oncogene)，突变体中则过度表达，等位基因中的单个突变可以影响细胞的表现型。癌基因发生突变的机制主要有：

(1)基因放大　许多癌细胞都含有多拷贝的正常癌基因，如乳腺癌通常含多量的 ERBB2 或 MYC，数量可达到比正常多数百个拷贝。可以以双微染色体(double minute chromosome)或插入正常的染色体中(同源性着色区，homogeneously staining region，

HSR)形式存在。

(2)点突变　　*HRAS* 基因是属于 *ras* 基因家族中,所编码的 p21 蛋白参与 G 蛋白偶联受体的信号传导(transduction)。从受体来的信号激发 GTP 与 RAS 蛋白的结合,再由 GTP-RAS 传出信号。RAS 蛋白 GTP 有酶活性,GTP RAS 迅速转换为没有活性的 GDP-RAS。在结肠癌、肺癌、乳腺癌和膀胱癌中通常都可以发现有 *ras* 基因的特殊点突变,导致某一氨基酸替换,降低 RAS 蛋白的 GTP 酶活性。结果 GTP-RAS 信号失活变慢,形成少量的受体信号产生广泛的细胞反应的局面。

(3)染色体移位产生新的嵌合基因　　目前已知有十余种与肿瘤有关的嵌合基因是由此产生的。其中最著名的是产生 Ph^1(Philadelphia)染色体的肿瘤特异性的重排,Ph^1 染色体见于 90%的慢性骨髓性白血病。该染色体是由一个平衡 9;22 交互移位所产生的。9 号染色体上的断点位于 *ABL* 癌基因的内显子中。移位把 ABL 序列的大部分加到一个称为 *BCR*(breakpoint cluster region)基因中(位于 22 号染色体),产生一个新的融合基因,该基因表达酪氨酸激酶,此酶与 ABL 产物有关,但具有异常的转变特征。

(4)通过转座到活性染色质区域而激活　　Burkitt's 淋巴瘤发生过程中 *myc* 癌基因的激活是关键的一环。在 75%～85%病人细胞内有特征性的 t(8;14)(q24;q32)染色体易位,其余的为 t(2;8)(p12;q24)或者 t(8;22)(q24;q11)易位。结果使 *myc* 基因靠近一个免疫球蛋白基因座,即 14q32 的 *IGH*、*IP*12 的 *IGK* 或 22q11 的 *IGL* 基因座。和嵌合基因不同,这种易位不产生嵌合基因,而是把癌基因放在了产生抗体的 B 细胞中的活性转录染色质区域。通常 *myc* 基因的外显子 1(非编码)不被同时易位。由于失去了正常的上游控制,并放到活性染色质区域中,使得 *myc* 于一异常高水平表达。这类癌基因激活主要是处于白细胞和淋巴细胞中,而不太可能在实体癌中发生。

2. 肿瘤抑制基因

肿瘤抑制基因产物抑制细胞增殖,癌细胞中的突变体丧失其功能。*TS* 基因的两个等位基因都必须失活以改变细胞的表现型。

肿瘤抑制基因发生突变时也可导致癌症(见表 13-11)。其在正常机体的功能很难检测到,有些功能较简单,如 APC 和 DCC 可能参与细胞黏合。其他 *TS* 基因产物可能与细胞周期循环有关,是作为负性调节因子。其中较典型的是 *RB* 和 *TP53*,这两种基因的产物在控制细胞周期进行中起重要作用,在肿瘤细胞中经常可见到这两种产物都已经失活。其机制见图 13-17。

表 13-11　*TS* 基因突变引起的罕见的家族性癌症

病　名	MIM 编号	染色体定位	基因
Von Hippel-Lindau 病	133300	3p25～p26	*VHL*
家族性腺瘤型结肠息肉	175100	5q21	*APC*
家族性黑素瘤	600160	9p21	Gorlin's *CDKN2*
基底神经节综合征	109400	9p22～q31	(*BCNS*)
多发性内分泌瘤 2	171400	10q11.2	*RET*

续表

病　名	*MIM* 编号	染色体定位	基因
Wilms 瘤	194070	11p13	*WT1*
多发性内分泌瘤 1	131100	11q13	(*MEN1*)
运动失调性毛细管扩张症	208900	11q22～q23	*ATM*
早发型乳腺癌	600185	13q12～q13	(*BRCA2*)
Li-Franmeni 综合征	151623	17p13	*TP53*
神经纤维瘤 1	162200	17q12～q22	*NF1*
乳腺-卵巢癌	113705	17q21	*BRCA1*
神经纤维瘤 2(双侧神经瘤)	101000	22q12.2	*NF2*

3. 增变基因

增变基因(mutator gene)负责维持染色体组的完整性和信息传递的保真度。两条等位基因的功能丧失时细胞则能在基因突变(如癌基因和 *TS* 基因发生突变)后仍发展。

增变基因与癌基因及肿瘤抑制基因不同,它不参与细胞周期的控制,而是在保证遗传信息的完整性中起作用。这类基因的突变导致 DNA 的不完全复制或修复。很早就知道癌细胞表现为遗传不稳定性。它们都有很不正常的核型,染色体有许多缺失、获得和重组,但其中仅有一少部分与癌症有关。目前这些负责染色体稳定性的基因还未鉴定出来。仅有结肠癌和运动失调性毛细血管扩张症中对 DNA 水平的不稳定性有关基因进行了研究。与结肠癌有关的增变基因见表 13－12。

表 13－12　结肠癌中的增变基因

大肠杆菌	人类	在人染色体中定位	HNPCC 百分率
MutS	*MSH2*	2p15～p22	50%～60%
MutL	*MLH1*	3p21.3	30%～40%
MutL	*PMS1*	2q31～q33	5%
MutL	*PMS2*	7p22	5%

注:HNPPC 为遗传性非息肉型结肠癌 (hereditary nonpolyposis colon cancer),目前癌症发生的详细过程还不太清楚,但从结-直肠癌的一些细节,可大致拼凑出各种与癌症有关的基因是如何参与癌症发生的(见图 13－17)。从中可见癌症的发生是一个多步骤的演化过程。

四、基因突变方法在研究中的应用

1. 离体定向诱变方法

这种方法是指人们有目的地在离体条件下制造位点特异性突变,然后设法导入生物体内,观察和分析这些突变的表现型,常称为反求遗传学 (revers genetics) 或 DNA 分析遗传学(genetics by DNA analysis)。这种方法在基因的表达和调控机制研究中起了很重要的作用。一般的离体定向诱变方法包括下面几个步骤:

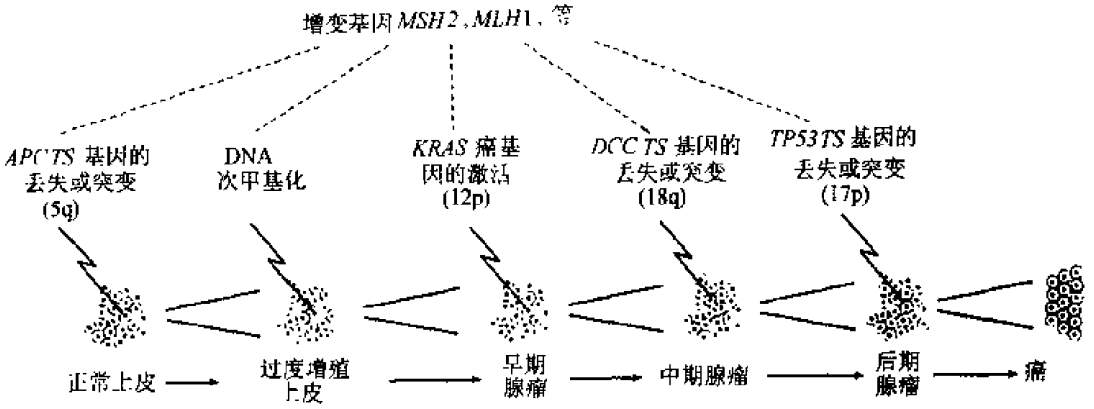


图 13-17 Feron 和 Vogelstein 的直肠癌发生模型

1）合成与野生型基因同源的但同时也有突变位点的寡聚核苷酸链；
2）分子杂交。即把合成的寡聚核苷酸和由载体携带的 DNA 克隆的互补单链混合进行杂交；
3）进行转化，由细菌内部的复制和修复机制得到稳定遗传的突变 DNA 克隆；
4）将突变基因送回细胞内通过等位基因代换进入目的基因组中；
5）突变检验，及其他性状分析。

2. 转基因动物方法

转基因动物技术是目前生命科学研究领域中最激动人心的进展之一。Gordon 于 1980 年将胸腺激酶基因和 SV40 病毒基因片段的融合基因用显微注射法导入小鼠胚胎细胞的原核中，获得了转基因小鼠。1981 年 Gordon 和 Ruddle 首创了“转基因”(transgenic)一词，意指动物携带有新的基因，并把外源性 DNA 片段整合到其基因组中。这种动物称为转基因动物。次年，Palmiter 等人用小鼠金属硫蛋白基因-1(MT-1)启动子片段与人生长激素基因构建的融合基因，成功地生产出了整合有人生长激素基因的转基因“巨型小鼠”，生长速度快，个体是正常小鼠的两倍以上。随后，生命科学的诸多领域都采用了转基因动物方法，进一步推进了其发展。

转基因动物方法是指将外源基因导入到动物的 1 细胞胚胎的原核中，该胚胎最终发育为携带有外源基因并使其整合在基因组上的新个体。因此转基因动物的生产是一个包括目的基因构建、纯化，供、受体动物的处理，DNA 的导入，胚胎的收集、培养及移植等极为复杂的系统工程。其中任何一个环节都直接影响转基因动物生产的效率。而 DNA 导入是诸环节中最重要的。目前 DNA 导入的方法有多种，常用的有显微注射法、反转录病毒法、胚胎干细胞法及精子载体法等。

转基因动物在生物医学、农业、生物学及生物工程等领域具有巨大的应用潜力。首先在人类疾病治疗中，转基因小鼠作为人类疾病的模型，使得人们几乎可以针对任何一种疾病设计相应的转基因小鼠模型，如癌症、代谢疾病、高血压、动脉硬化、中枢神经系统免疫性疾病和退行性疾病等。许多疾病都是一种或几种蛋白表达量不正常所致，如将有关基因分离出来并克隆，制备相应的转基因动物，就可以对这些疾病进行深入的研究，为人类

疾病的病因、诊断和治疗提供可靠的依据。现已建立了生长病、免疫病、神经病、生殖病及血液病等模型。其次，在生产具有医药价值的稀有生物活性蛋白中具有特殊价值。将编码具有某种活性的蛋白质的基因制备转基因动物，使之在动物的血液或乳汁中表达，即动物生物反应器。现已在小鼠、家兔、猪和山羊上获得成功，生产包括人血红蛋白、α-抗胰蛋白酶、凝血因子 IV、组织纤维蛋白溶解酶原激活剂(tPA)、尿激酶(UK)、尿激酶原(Pro-UK)、C-蛋白、促红细胞生长素(EPO)等。上述物质多数以极微量水平存在于人的血液或组织，但具有极高的生物学活性，如果用分离提取的方法产量太少，体外转染微生物生产这些药物性蛋白无法完成转录后的恢复，体外动物细胞培养费用昂贵。而转基因动物恰好解决了上述难题，并可以每升几十克的水平在其乳汁中生产药物性蛋白。在畜牧业改良和培育动物新品种方面，转基因技术则克服了原先培育速度慢和亲缘关系的限制，使新品种的培育大大加快。可以培育提高饲料转化率和生长速度，减少胴体脂肪，增加畜产品如毛、乳、肉的产量和品质，如在反刍动物的乳腺生产拟人化的乳汁——含有人的乳铁蛋白，具有强的抗菌特性，并在哺乳婴儿的离子吸收方面起作用。

然而，转基因动物仍然存在许多问题有待解决，如转基因效率低下，转基因的随机整合、不表达或异常表达，转基因动物生产性能低下和出现病态等等。但这些问题都不能掩盖转基因方法在各个领域所产生的深远影响。

(阮承迈)

参 考 文 献

[1] Luria S E and Human M L. J Bacteriol, 1952. 64: 557~569
[2] Betani G and Weigle J J. J Bacteriol, 1953. 65: 113~121
[3] Arber W and Dussoix D. J Mol Biol, 1962. 5: 18~36
[4] Arber W and Dussoix D J Mol Biol, 1962. 5: 37~49
[5] Arber W. J Mol Biol, 1965. 11: 247~256
[6] Yuan R. Ann Rev Biochem, 1981. 50: 285 -315
[7] Little J W and Mount D W. Cells, 1982 29: 11~22
[8] Walker G C. Microbiol Rev, 1984. 48: 60~93
[9] Marinus M G. Ann Rev Genet, 1987. 21: 113~131
[10] Lindahl T et al Ann Rev Biochem, 1988. 57: 133- 157
[11] Sancar A and Sancar G B. Ann Rev Biochem, 1988 57: 29~67
[12] Prakash S et al. Ann Rev Genet, 1993. 27: 33~70
[13] Bickle T A and Kruger D H. Microbiol Rev. 1993. 57: 434~450
[14] Friedberg E C, Walker G C and Siede W. DNA Repair and Mutagenesis. American Society for Microbiology, (Washington DC). 1995
[15] Boyer H W and Roulland-Dussoix D. J Mol Biol, 1969. 41: 459~472
[16] Yuan R and Meselson M. Proc Nat Acad Sci USA, 1970. 65: 357~362
[17] Hadi S et al. J Mol Biol, 1979. 134: 655~666
[18] Yuan R. Cell. 1980, 20: 237~244
[19] Fuller-Pace F V and Murray N E. Proc Nat Acad Sci USA, 1986. 83: 9683~9372
[20] Klimasauskas S et al. Cell, 1994. 76: 357~369
[21] Sancar A and Rupp W D. Cell, 1983. 33: 249~260
[22] Sancar A. Ann Rev Biochem, 1996. 65: 43~82
[23] Wood R D. Ann Rev Biochem, 1996. 65: 135~168
[24] Cooper DN, Krawczak M. Human Gene Mutation, 2nd Eds. BIOS Scientific Publishers, Oxford. 1995

[25] Beaudet A, Tsui LC. Hum Mutat, 1993. 2, 645
[26] Strachan T, Andrew PR. Human Molecular Genetics, BIOS Scientific Publishers. 1996
[27] Lewin B. Genes. Oxford University Press. 1997

第十四章　遗传与进化

进化论是整个生物学的指导思想。杜布赞斯基(Th. Dobhansky)说:"没有进化论,生物学便成为不可理解的"。因为,一大堆割裂开的事实、片段,在没有把它们用一根线串联起来成为一个整体观念之前,决不是科学。这根赋予生物学以生命的线,就是进化论。进化论的学科研究从达尔文(Darwin)和华莱士(Wallace)1858年发表的论文开始的,他们首次提出生物进化是自然选择的结果。只有存在遗传变异时自然选择才是有效的,而遗传变异主要是由突变和重组提供的。在达尔文时代,还不了解遗传变异的产生机制,孟德尔(Gregor Mendel 1822～1884)遗传定律的发现,为遗传学的发展殿定了基础,该定律被发现于1865年,但被埋没了35年后才被人们了解。随着现代遗传学和分子进化论的发展,遗传科学本身发生了革命,这个革命始于1953年,沃森(J. Watson)和克里克(F. Crick)发现了DNA分子的双螺旋结构,基因本身——DNA的分子片段——某种有机化合物(多肽)的成分记码,多肽形成酶和其他决定有机体从受精卵发育的物质。从此,明确了许多基因活性的本质及其控制的问题,为我们提供了一个新的分子遗传学的范例,它为进化论研究者提供了新的研究进化的方法,在分子水平上直接研究进化的过程——基因及其产物。近年来由于重组DNA技术、聚合酶链式反应(polymerase chain reaction, PCR)、DNA限制性片段长度多态性(restriction fragment length polymorphism, RFLP)和随机扩增多态DNA(randomly amplified polymorphic DNA, RAPD)方法的开发应用,使遗传与进化的研究发展日新月异,从而从分子水平上解释遗传现象。DNA的核苷酸序列中储存的遗传信息,基因组的结构、复制、转换、修复、变异以及控制基因表达的各个元件等都是其主要的研究内容。

第一节　遗传变异与进化[1～9]

当人们说起生物进化时,他们想到的总是通过很长时间的逐渐改变,某些种类的有机体转变成为另一些种类。进化的基础是遗传的改变,进化的发生是因为有机体的遗传结构由一代传到另一代发生了变化,也就是说进化是一个由代代相传的遗传变化构成的积累过程。遗传变异的存在是进化的必需条件。如果一群体的所有个体在某个基因位点的遗传性是等同的,就是说确实是同合的,进化就不可能在此位点上发生。因为等位的频率代与代之间不可能改变。另一方面,如果在另一群体,同一基因位点有两个不同的等位,当一个等位的频率增加,另一个等位的频率有所减少时,在这个位点上就可能发生进化。

一、遗传变异的起源

生物学的常识告诉我们：最原始的生命是非常简单的和非常小的。然而一切物种皆从这些最简单的生物开始，现在的生物已达两百多万种。它们在大小、形状和生活方式上非常多样化。现代生物学知识告诉我们：带有遗传信息的 DNA 序列也一样存在多样化。在进化过程中，已有的 DNA 序列一定要经历改变，新的序列必须加入到生物的基因组中去，才可能使生物进化发展。遗传是个保守的过程，但也并非完全如此，否则进化就不可能发生了。在 DNA 的核苷酸序列中记录的信息，作为一条规律，是忠实复制的，所以，每次复制的结果是形成两个 DNA 分子，彼此等同和它们的亲本也等同。然而，DNA 分子在复制过程中也会发生变化，导致 DNA 分子产生不同的核苷酸序列，或者在亲子细胞之间，其 DNA 数量上产生差异。这种在分子水平上形成的遗传材料的改变叫突变。它是遗传变异产生的内在根据。突变为进化提供了最根本、最原始的材料。突变可分为基因突变和染色体畸变。

二、基因突变与进化

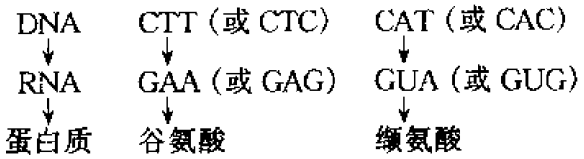
基因突变(gene mutation)是指染色体上一个基因座位内遗传物质的改变，由此产生出等位基因(allele)或复等位基因。

基因突变现象最早由 DeVries 于 19 世纪下半叶在一种植物月见草中发现的。1909 年 Morgan 在红眼果蝇群体中发现了白眼果蝇，后来研究证明白眼性状是由于 X 染色体上的红眼基因(W)发生突变($W \rightarrow w$)所引起的。随后研究发现果蝇存在几百种突变。根据大量研究的事实逐渐明确动植物及微生物中很多单位性状内的差别都来自该生物进化过程中的基因突变。例如水稻由非糯性变为糯性($Wx \rightarrow wx$)，小麦由高秆变为矮秆($D \rightarrow d$)等等，这些性状的改变都是基因突变的结果。

由于对遗传物质 DNA 的性质和功能的了解，人们认识了各种基因。结构基因编码 mRNA，进而决定蛋白质合成；另一些基因编码 tRNA；还有一些基因编码 rRNA。看来，有很大部分 DNA 执行各种调节功能，高等生物尤其如此。突变大概可以影响这些类型的任何基因，但一般研究过的是结构基因突变。基因突变就是某一遗传密码中一个或几个核苷酸(碱基)的替换，以及核苷酸的增加或丢失(见第十三章)。

基因突变的效应是多种多样的。多数突变是微效和有害的。不错，确有一部分突变是中性的。突变是否改变了表现型和改变多少一般要看所产生的多肽链(蛋白质)的功能是否发生了改变和改变多少。现在知道，基因突变从没有效应到有害效应和有利效应，各种程度都有，情况相当复杂，最严重的是致死效应。

人类的镰状细胞贫血症是一种隐性致死的遗传病，纯型合子使红细胞呈镰刀状，引起严重贫血而致死。据分析，这是由于组成血红蛋白的β链第 6 个氨基酸的改变，即谷氨酸被缬氨酸所代替。这是有关遗传密码的一个碱基的代替：



从上面的材料可以看出，如果遗传密码 CTT 的第三个碱基被碱基 C 所代替，遗传密码成为 CTC，那么效应并没有改变，因为这两个遗传密码通过转录，都会翻译成谷氨酸，这属于中性突变。

三、染色体畸变与进化

染色体畸变（chromosome aberration）跟基因突变一样，在生物进化中占有重要的位置。基因突变和染色体畸变所产生的变异都是可遗传的。两者在发生机制中有明显的区别。基因突变大都是 DNA 分子中某个遗传密码的一个或几个碱基的变化。染色体畸变则牵涉到 DNA 分子在较大范围内的变化。例如，染色体片段或整个染色体的增加、丢失或者改变位置（易位或倒位）等。

基因突变和染色体畸变都是有关遗传物质的变化。基因突变是个别基因分子基础的化学变化，例如，一个碱基被另一个碱基所代替，或者增加了个别碱基或者减少了个别碱基等。染色体畸变通常并没有发生什么深刻的化学变化，而是已有的遗传物质或已有的基因排列顺序进行了重新改变，可是后果却很严重。

引起基因突变的因子通常也可以引起染色体畸变。

基因突变一般发生在遗传物质复制的过程中，而染色体畸变通常发生在细胞核分裂的过程中，特别是减数分裂的过程中。有资料表明，减数分裂的前期染色体比较敏感，比较容易发生断裂。

染色体畸变可包括两个主要方面：一是染色体数目的异常；二是染色体结构的异常（见第一章）。

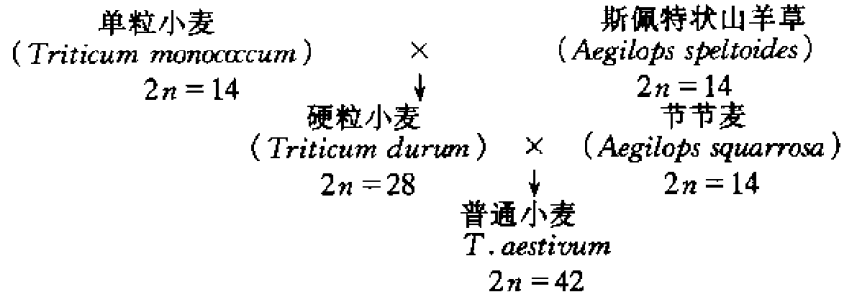
1. 染色体数目异常与进化

染色体作为一个结构单位，为保持个体和群体在遗传上的连续性和稳定性，提供了物质基础。由于它在每一世代中都以十分精确的程度进行自我复制，使物种得以延续。然而生物的多样性是生物界的明显特征，而生物的进化则是多样性的源泉，它是生物个体在遗传的基础上其遗传物质逐代改变的结果。也就是说，细胞中的遗传信息保守地传递给后代，但也会发生变异，造成染色体结构和数目的变异，同时也引起其所载荷的 DNA 序列和数量的改变，并在自然选择的基础上，进行新性状的积累。由此说明，遗传使物种得以保存，性状得以继承，而变异正是对这种遗传稳定性的动摇和突破，从而产生新物种。

染色体数目异常主要包括整倍体和非整倍体。

例如，在植物界，多倍体物种很普遍。在开花的种子植物中，几乎有半数的物种是多倍体。在动物界，多倍体物种也有一些，但是比较少见。

普通小麦（*Triticum aestivum*）的染色体是 $2n=42$，这实际上是一个六倍体物种。它的起源大抵如下：



在上述的小麦进化中，杂种的染色体如何加倍，这是历史上发生的事。现在在实验室里，使远缘杂种的染色体加倍，可以利用秋水仙素。利用这种物质，可以使小麦和黑麦杂交产生的杂种染色体加倍，成为能育的物种——小黑麦。

非整倍体异常的典型例子是人的 Down 氏综合征(Down syndrome)，即先天愚型，是由 21 号染色体三体引起的，男性和女性都有。女性是：47，XX，+21；男性是：47，XY，+21。这是人类中最常见的常染色体异常。这和人类及其他物种中三体的不利作用很一致。因此，染色体数目的增加对发育也有不利的影响，这可能是因为多余的染色体打乱正常基因组赋予的发育平衡的缘故。因此，非整倍性在动植物进化中没有起什么大的作用。多体性是额外的遗传物质能够整合进基因组的一种机制，但证据表明，这种增加并不普遍。这不是说基本的单倍体数目的变化没有发生，而是说单倍体数目的增减通常不是由整条染色体的简单增加和丢失造成的。

2. 染色体结构异常与进化

这是染色体在内部结构上的重新排列，是由于一条或一条以上的染色体发生了横的断裂，随后在断裂口反常地重新连接起来，于是遗传物质原来的直线排列就发生了紊乱。由此产生若干类型的染色体结构异常。主要有倒位(reversion)、易位(translocation)、缺失(deletion)和重复(duplication)。

倒位　只改变基因在染色体中的顺序，不改变基因原来的数目，因此，倒位纯合体及其配子有正常生活力，其自交后代一般表现完全正常。但由于倒位区段内的基因与连锁群中其他基因的交换值有了改变，影响其杂交后代的重组率，同时可能引起基因的位置效应。倒位杂合体其倒位圈内连锁基因的交换可能受到抑制，因此常会降低正常交换值。另外，倒位圈内的基因若发生交换，将形成一部分缺失某些基因，重复另一些基因的配子，这些交换配子没有生活力，因此也造成了后代重组率的下降及高度不育。此外，由于染色体可能多次在不同世代中发生倒位，加上自交的作用，产生的倒位纯合体后代将逐渐与原来物种存在生殖隔离而形成新物种或变种，可见倒位也可以促进生物进化。根据对果蝇唾液腺染色体的研究，发现果蝇属中存在着不同倒位特点的种，例如 *D. melanogaster* 和 *D. simulans* 是两个比较相似的种，它们之间的杂种一代完全不孕。检查杂种一代的唾液腺染色体，发现在一条染色体上只有一个大的倒位，其他染色体在基因排列上有几处小的差异。

易位　在遗传及生物进化上也有重要意义。一般来讲，易位并未改变染色体和基因的数目，只改变了其原来的位置。易位也能改变基因连锁群，原来有连锁关系的基因易位后可能成为独立遗传关系的基因；原来独立遗传的基因，易位后反而可能建立连锁关

系。易位也同样可能产生位置效应，从而引起表现型的某些变化。现已证明，很多植物的变种就是染色体在进化中一再发生易位而形成的。在家马中，所有品种和类型的马都是64条染色体($2n$)，然而被认为是家马祖先的蒙古野马(*Equus prezewals* Rii)，其染色体数目是66条。据分析，这种野马与家马相比，野马少两条中央着丝粒染色体，而多4条近端着丝粒染色体，因此有些学者认为家马的这两条中央着丝粒染色体是由野马中4条近端着丝粒染色体通过罗宾逊易位方式而形成的。染色体易位不仅在理论上具有一定的意义，在生产实践中也具有重要价值。如在养蚕业中，雄蚕的吐丝量比雌蚕高20%～30%，而且质量好。如果多养雄蚕就会带来更大的经济效益，但在大规模生产中逐个鉴别雌雄是不可能的。因此通过人工易位的方法，将黑色卵色基因易位于W染色体上，从而成功地培育出雌雄蚕卵自动鉴别品系，大大提高了蚕丝的产量和质量。

缺失(deletion)和重复(duplication)　这是同源染色体之间发生非均等互换的结果。一条染色体得到较多的染色体片段，另一条染色体得到较少的染色体片段。片段的缺失造成部分遗传物质的缺失。缺失纯合时往往致死，若缺失大到一定程度，在杂合体中可引起显性致死作用。因此，缺失在进化中往往不起重要作用。

重复对适合度的影响通常不如缺失的影响严重，它长期被看作是额外遗传物质加到基因组的一种机制。曾经设想，原始基因的功能分化是通过突变而发生的，因为继续执行原有功能的某个基因的存在，使重复基因不受约束，容易发生突变，执行新的功能。由于进化的总趋势是增加DNA量和功能更加复杂化，所以，人们认为重复有很大的进化意义。因为重复的基因通过一系列的突变可以形成一个新的非等位基因。这是新基因起源的重要途径之一。例如，重复了的基因A，就在同一染色体上合成为AA。于是在二倍体细胞里就成AA/AA。其中的一个A如果发生突变，功能也改变了，那么新基因(叫做B)就出现了。由于A还存在，并不影响个体原来的代谢功能。但是由于其中的一个$A \rightarrow B$，即会出现新的基因和新的功能。

第二节　DNA的进化变异[10~24]

DNA不仅是主要的遗传物质，同时，也是生物进化史的重要记录者。通过DNA序列分析研究生物的进化历程。确定物种间的进化关系具有许多优越性。

一、DNA的含量与进化机制

在生物进化过程中，DNA的含量明显增加(图14-1)。虽然现在的病毒并不能代表最原始的生物，但在$\Phi X174$和F1等病毒中仅有6～8个基因的DNA含量(约相当于6000个核苷酸长)。而哺乳动物的每一个单倍染色体组就有3×10^9个核苷酸对，假如全部DNA都能传递遗传信息，它就相当于300万个基因。这种DNA含量的增加，在生物由较简单类型向复杂类型的进化中，显然是重要的。复杂的生物要维持生命，就必须具备大量的基因。实际上，很多基因仅高等生物才有，例如，血红蛋白基因、免疫球蛋白基因等就是。

但是，最近研究表明，DNA含量与生物形式的复杂程度之间并不一定有相关性

(表 14-1)。

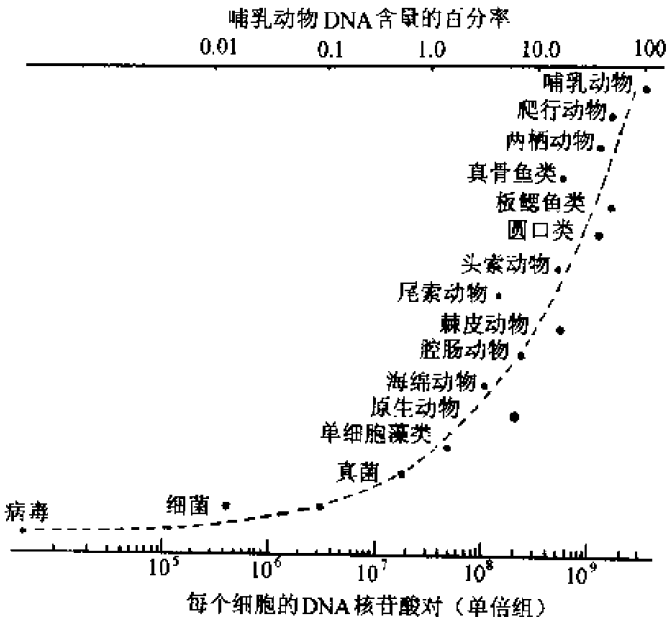


图 14 1 不同生物物种的 DNA 含量的最低限度值

图中各圆点表示一个单倍染色体组每个细胞的 DNA 含量测定值。

例如，有一种肺鱼的 DNA 含量比哺乳动物高 40 倍。许多两栖类的 DNA 含量也远远超过哺乳动物种。可见，单凭较高的 DNA 含量还不足以产生一种复杂的生物，而需要基因组中拥有足够数量的各种基因。

表 14-1 不同生物的 DNA 含量

生物	每个染色体组的核苷酸对数	生物	每个染色体组的核苷酸对数
哺乳类	3.2×10^9	果蝇	0.1×10^9
鸟类	1.2×10^9	玉米	7×10^9
蜥蜴类	1.9×10^9	链孢霉	4×10^7
蛙类	6.2×10^9	大肠杆菌	4×10^6
大多数多骨鱼	0.9×10^9	T4 噬菌体	2×10^5
肺鱼	111.7×10^9	λ 噬菌体	1×10^5
棘皮动物	0.8×10^9	ΦX174	6×10^3

高等生物的 DNA 含量较高，主要是进化过程中基因重复的结果。基因重复有两种类型，一种是染色体重复，另一种是不等交换产生的染色体小片段重复(衔接重复)。染色体重复的常见形式是基因组重复。正如 Stebbins 指出，基因组重复或多倍化在植物进化中起重要作用。在动物中，以往一直推想不等交换是遗传物质增加的主要机制。不过，关于核仁 DNA 含量的最新研究表明，不同生物之间核仁 DNA 含量的变异很不一致。因此，基因组重复在动物进化上也是相当重要的。Ohno 通过细胞学和生物化学研究推断，在大约 3 亿年以前的鱼类阶段，哺乳动物谱系中就至少发生过一次多倍化过程。并指出在性染色体分化之前，基因组重复在动物进化中就是经常发生的。一旦哺乳类、鸟类及爬行类谱系中性染色体完成分化，基因组重复就会破坏性别决定机制，而几乎立即停止形成

四倍体。在大部分鱼类和两栖类中，性染色体尚未分化，因而四倍体雄性和雌性较易保存下来。

在有机进化中，不等交换产生的衔接重复显然具有同等重要的作用。支配相同或相似功能的基因常常是紧密连锁的。例如，普通果蝇的每条 X 染色体和 Y 染色体的核仁组成中心区段内，就集结了大约 100 个核糖体 RNA 重复基因。同样，编码数种免疫球蛋白多肽的同源基因也是紧密连锁的。至于这些紧密连锁的同源基因的进化，用衔接重复可以得到满意的解释。Horowitz 指出，伴有了基因功能逐渐分化的反复衔接重复的过程，也可以产生出细菌的操纵子，不过此时操纵子的结构基因仍然是同源基因。

统观真核生物细胞的 DNA 碱基序列，可以分为两大类型。一是不重复的序列。这是基因的编码基础。二是重复的序列，这包括以下的成分：①重复的 rDNA、tDNA，它们分别转录为有关的 RNA；②重复的组蛋白 DNA，它通过转录和翻译，产生出组蛋白；③不含有编码意义的序列，这可以重复几次到几千次或更多；④不含有编码意义的序列，这见于割裂基因(split genes)中的内含子(intron)；⑤不含有编码意义序列，这见于染色体的卫星(satellite)DNA 中。

突变可以发生在有编码意义的序列中，也可以发生在没有编码意义的序列中。从现在已知的材料看，DNA 的分子进化与蛋白质的分子进化有相关性。但是，DNA 有自己的进化史，因为，并不是 DNA 分子中的序列都有编码的意义。

二、DNA 重复序列的进化

原核生物一般只有一个染色体即一个核酸分子(DNA、RNA)。而真核生物的染色体中，DNA 约占 27%，组蛋白和非组蛋白占 66%，RNA 占 6%。由于真核生物的 DNA 比原核生物的大得多，DNA 序列的组织性也就复杂得多。近年来发现许多 DNA 序列起源于重复，有时这种重复序列再进一步分化。一切基因必定都是由单个或少数几个原来的基因重复而来。可以作为重复而识别出来的 DNA 序列有：

(1)轻度重复序列　在一个基因组中有 210 个拷贝，如组蛋白基因和酵母 tRNA 基因。在某些情况下，这些序列并非一点不差的重复，这些基因的产物在氨基酸组成上略有差异，如胚胎、胎儿、成人的血红蛋白 β 链的一系列基因就是如此。

肌红蛋白和血红蛋白是脊椎动物红血球的主要成分，其功能是运输氧气和二氧化碳。

肌红蛋白具有单个多肽链，有很复杂的三维空间结构，分子量约 17 000 Da。血红蛋白一般含有四个亚单位，两个亚单位属一类，另两个属另一类，其分子量约为 67 000 Da。在人类成体中，有两类血红蛋白，它们是 A 和 A2。血红蛋白 A 最为普遍，由两个 α 和两个 β 多肽链组成。还有 α2β2 组成的血红蛋白 A2，在成体中血红蛋白只有 2% 含有两个 α 和两个 δ 链者。在人类胚胎中还发现第三种血红蛋白，叫做胎儿血红蛋白，具有两个 α 和两个 γ 链。α、β、δ、γ 多肽链各自有不同的基因记码。Ingram 就是通过测定人类血红蛋白的 α 链、β 链、γ 链以及肌红蛋白的氨基酸序列，推断人类血红蛋白三条链的基因是基因重复产生的。这三条链的比较表明，α 链与 β 链的共同氨基酸比率达到 41%，而 β 链与 γ 链的共同氨基酸比率更高，为 73%。由于这三条链高度相似，因此这种相似性的随机出现概率可以忽略不计。Ingram 又指出，人的肌红蛋白同血红蛋白的三条链一样，也是起源

于同一祖先。

(2)*中度重复序列*　10至几百个拷贝。中度重复序列一般是不编码的序列，据认为在基因调控中起重要作用，包括开启或关闭基因的活性，促进或终止转录，DNA复制的起始，其转录产物参与hnRNA的处理。这种重复序列的平均长度为300个bp，常有数千种这样的重复序列，每一种平均重复数百次。它们在一起构成了序列家族。它们与非重复序列相同排列。其中最使人感兴趣的是人类的*Alu*序列家族(*Alu* family)。例如，位于人的甲状旁腺激素基因的5′侧翼的负钙应答因子已被证明位于一*Alu*序列内。这个基因由一个回纹核心序列和几个上游的T残基组成。*Alu*序列在转录起始位点的上游316 kb左右的位置，以反向组分出现，并以一个长的polyT序列开始。polyT和相邻的14个核苷酸形成负的钙应答因子位点。

(3)*高度重复序列*　几百个到几百万个拷贝。在高度重复序列中，常有一些A、T含量很高的简单高度重复序列，例如螃蟹DNA含有AT的简单重复序列，有时在30个左右的碱基对中才偶尔插入一个G、C对，因而A、T含量达到97%。由于A、T的段浮力密度较小，因而在将DNA切断成数百个碱基对的片段进行超速离心时，常会在主要的DNA带的上面有一个次要的DNA带相伴随，这就是所谓卫星DNA(主带的AT含量为58%)。那么，卫星DNA是如何起源和进化的呢?

三、卫星DNA的等级结构及其起源和进化

哺乳动物的卫星DNA常常是多等级的，就是说一个大的重复单位是由若干个彼此相似的小重复单位串联组成；每个小重复单位是由若干个彼此相似的更小的重复单位串联组成的。例如将小鼠卫星DNA用限制酶*Eco*RII进行切割，可以得到230～240 bp的片段，通过序列分析发现在一个234 bp序列中，60%～70%的卫星DNA都含有这样的序列。如果把234 bp的左一半和右一半各117 bp排起来，发现只有22个bp不同，差异为19%。这说明现在的234 bp重复单位是由117 bp亚重复单位倍生而来，然后由于突变而积累了差异。117 bp的亚单位又可分为58 bp的四分之一亚单位。4个58 bp亚单位之间的差异达到40%。58 bp的亚单位又可以分成两个，一是28 bp(α亚单位)，一是30 bp(β亚单位)。8个八分之一亚单位之间的差异达61%。如果再将八分之一亚单位分成3个部分，则每一部分都与如下一致序列相近似。

GAAAAACGT
GAAAAATGA
GAAAAAACT

小鼠的234 bp重复DNA序列可能是起源于一个9 bp的序列(GAAAAATGT)，产生这一模型的一般原则是：在某一特定时间内由于某种原因一个重复单位突然横向扩增，以至产生多个相同串联重复拷贝。这种突然的横向扩增就称为跃进式复制(saltatory replication)。随后，由于突变的积累，这些拷贝发生分化。然后这些拷贝再经历一次跃进式复制，出现更长的重复单位；由于再经过突变的积累，使这更长的重复单位又发生分化。假定现在的卫星DNA起源于这样一个GAAAAATGT串联重复序列。最初的9个bp是相同的，随着时间的推移，突变使它们产生了差异。三个相邻单位的扩增使9 bp重复变成

27 bp 重复(假定为祖先序列)。在由 27 bp→58 bp 的过程中,还包括了两个插入突变,一个插入 C 成 α 亚单位,一个插入 3 个碱基成为 β 亚单位。将这 58 bp 描述成$(\alpha\beta)_n$。随着点突变、缺失和插入,重复单位的分化越来越严重。经过再一次的跃进式复制,58 bp(αβ)变成 116 bp(αcβcαdβd)的重复单位。随着进一步突变,两个相邻的 116 bp 被扩增成现在的卫星 DNA(见图 14-2)。

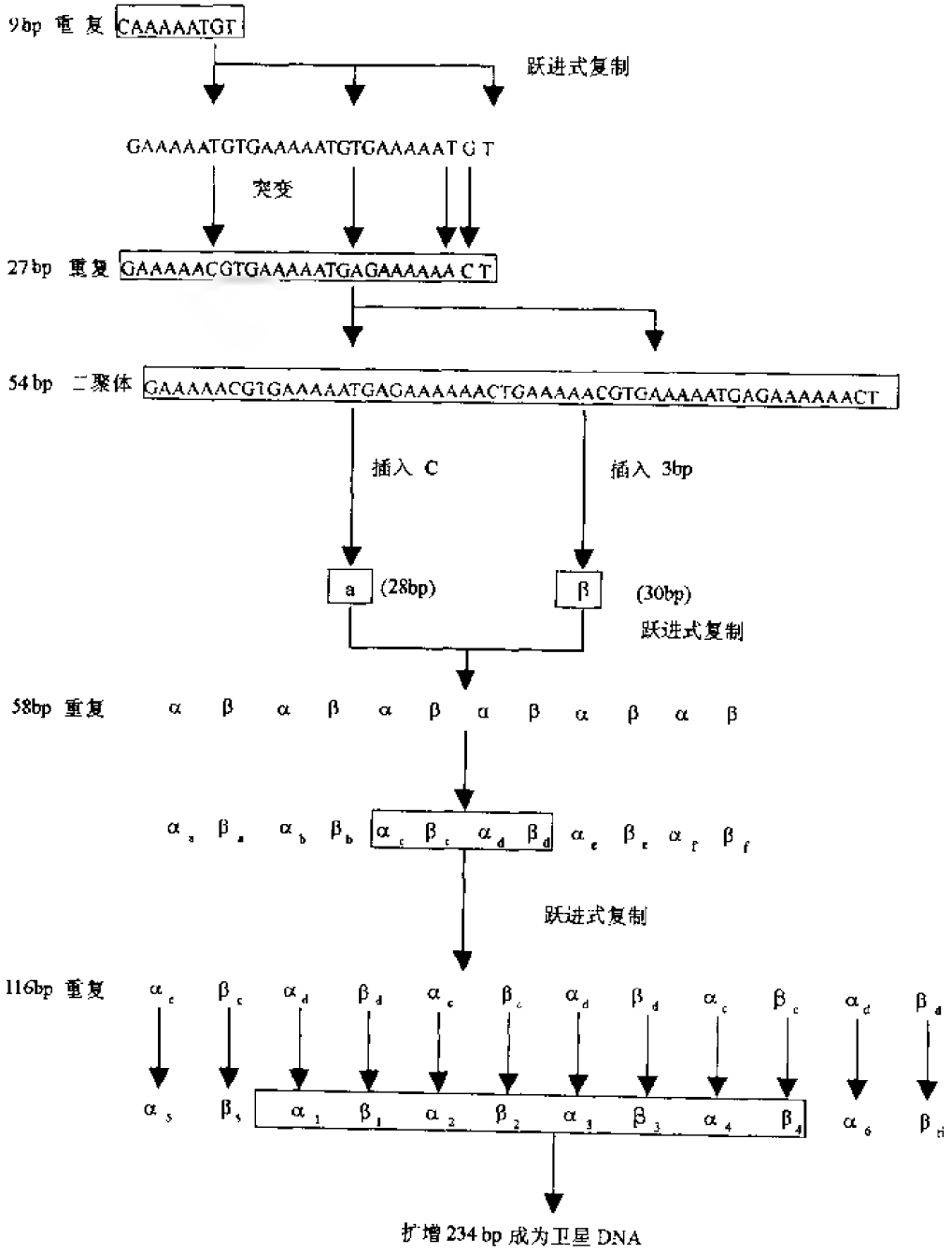


图 14-2 小鼠卫星 DNA 通过跃进式复制和突变积累的交替演化过程

高度重复 DNA 拷贝数的扩增或减少可能是由 DNA 的不均等交换引起的。细胞分裂时，染色体的同源重组是十分严格的，只能在同源序列之间进行。然而，由于卫星 DNA 的高度串联重复，一个重复序列很可能和另一个 DNA 分子上不等位的相同重复序列配对并发生交换。假如在两条染色体上各有 5 个 ab 串联重复单位，如果按图 14－3 发生非等位配对并发生交换时，则可得到一条含 7 个 ab 重复单位，而另一条只含 3 个 ab 重复单位。

图 14－3　DNA 的全符合状态的不均等交换

图 14－3 所示的不均等交换是全符合状态的(in register)，即某个染色体上的重复单位与另一条染色体上不等位的整个重复单位完全相配而发生不均等交换。如果在一个重复单位中的两个部分，如图 14-3 中的 a 和 b，又有很大的同源性，则 a 和 b 能发生配对而出现不均等交换。这种非等位配对是半符合状态(half-register)。这种不均等交换的结果如图 14-4 所示。

图 14－4　半符合状态的不均等交换

一般说来，卫星 DNA 没有受到任何选择压力，由于它不编码蛋白质或 RNA，其功能在于它的存在而不是其序列。这正如小鼠卫星 DNA 一样，其各个拷贝之间相关而不是相同。然而，有些卫星 DNA，如节肢动物的卫星 DNA 是由几乎相同的重复单位组成的。人类的 U2 snRNA 基因有 10～20 拷贝，在 17 号染色体上串联集中分布。每个重复单位约 6000 bp，然而其编码序列却只有 188 bp。编码序列受到自然选择的压力而保持相同，这是不难理解的。难以理解的是其余 5800 bp 的间隔序列为什么也要保持相同。当然其中包含了一部分转录起始、终止以及转录后处理的信号，它们必须是保守的。但是这一部分序列也只有 100 bp 左右，仍然无法解释其余 5700 bp 的保守性。从上面的例子可以得出以下结论：任何串联重复的 DNA 序列，不管其中是否含有编码的遗传信息，都将经受均一化作用(homogenization)。尽管这种均一化机制众说纷云，但其存在却是公认的。使串联重复的 DNA 序列保持均一化，目前，一般认为有两种机制：一是交换固定(crossover fixation)，二是基因转换(gene conversion)。这两种机制与基因扩增一起维持串联重复序列的均一化。

交换固定是由同源重组而引起的。也就是说在姐妹染色体上的非等位重复序列之间发生了同源重组，其结果产生了非均等交换。对于无编码功能的串联重复序列来说，非均等交换的两种产物均可能传递下去。因此突变序列或者被淘汰，或者迅速占据整个重复序列。而对于串联重复基因来说，大多数突变基因将被淘汰；但有功能的突变基因序列也有可能迅速扩展。在非均等交换中拷贝数减少了的重复序列可以在后代中通过基因扩增恢复原有的拷贝数。

基因转换也是通过同源序列之间的重组实现的。同源重组的中间体经拆分后形成异源双链，再经修复后产生基因转换。

四、“无功能”DNA

大部分高度重复 DNA 显然是无功能 DNA，它们不转录任何 RNA。一部分重复基因是无功能基因，可以用有害突变的积累来解释。Haldane 于 1933 年首次指出，假如基因组内存在两个或两个以上相同的基因，则除那些能产生必要数量基因产物的基因外，其余都是无功能基因。由于高等动物的进化过程中已发生过多次基因重复，所以很可能已积累了大量的无功能基因。Ohta 和 Kimura 根据遗传负荷假说，推算出哺乳动物的染色体组的 DNA 中有 90% 以上是无功能 DNA。Crick 推测，果蝇的结构基因分布在相当于整个基因组的 5% 的唾液腺染色体的间纹区段。但是，Turner 和 Laird 通过 RNA-DNA 杂交试验得知，全部 DNA 中至少有 24% 的 DNA 是能转录的。高等生物中功能性 DNA 的确切比例还有待进一步测定。

当重复位点中的某一位点出现致死的或无功能的突变时，只要其他重复基因或基因组保持正常的功能，则该突变将是无害的而且具有类似中性或近中性基因的作用。因此，在小群体内这种突变的固定率等于突变率。由于每个位点的每代致死突变率是 10^{-5}，如果存在许多重复基因，就会有大量基因变成无功能基因。当群体很大时，上述推论就不适用。在进化过程中，某些重复基因确实可以通过突变而获得新的功能。但是，突变是一种随机过程，所以发生上述事件的概率似乎是很小的。

假如上述假说是正确的，那又为什么会有这样多的 rRNA 或 tRNA 的功能性重复基因呢？这可能是由于蛋白质合成需要大量的 rRNA 和 tRNA 的缘故。如果这些位点中有多个位点发生致死突变，它们将降低杂合体的适合度，以致这些突变基因很快在群体中消失。实际上，如果无功能基因使杂合体适合度降到较低的水平，那么这些基因在重复基因位点上的固定概率就会明显下降。

人们知道，除了性决定基因、雄性能育基因以及核糖体 RNA 基因等一些特殊基因外，大多数生物的 Y 染色体缺乏功能基因。Muller 于 1914 年首次推测，Y 染色体的失活可能是致死基因积累的结果。他推断 Y 染色体上的基因位点经常保持杂合性，因此这些位点上发生的任何致死突变往往为 X 染色体的纯合位点上的野生型等位基因所掩蔽，而 X 染色体上出现的致死突变在同配性别中由于纯合性而消失。Muller 关于致死基因在掩蔽染色体积累的假说，也适用于二倍体或多倍体的无性生殖和单性生殖生物的染色体。由于这些生物不会发生分离和重组，因此，一个位点上的全部基因除一个之外都可成为无功能基因。掩蔽染色体的又一例子是始终保持杂合性的月见草的易位染色体。这种生物已积累了若干致死基因，所以易位纯合体不能存活。

五、DNA 序列插入和基因调控的进化变异

重复序列被认为是在基因的 5′区端，现已证明重复序列可以移动，并提供进化变异。最近的观察显示许多序列（已知的可移动因子、重复序列）被频繁地插入基因 DNA，而且

可以影响基因的表达调控。许多情况下，插入片段包括结合核蛋白的序列或插入后序列被修饰，结果形成结合位点，并成为基因调控系统的有效部分。有证据表明插入片段出现在好久以前，并发现在基因调控中 DNA 序列因子的插入是导致进化变异的根源，这种改变存在于长期的进化过程中。因而这个机制可被认为是形成基因调控区的 DNA 序列的来源之一。随着大量资料的获得，人们已知道这样一个事实，即一个插入子提供了新的调控因子。有关调控的进化起源知之甚少，序列改变的出现已有很长的历史，但它已掩盖了原来的结构，重复序列参与基因调控必须具备以下 4 个条件：①有出现在基因区的一个已知种类的因子的痕迹；②有证据表明该序列已存在足够长的时间，而不是瞬时突变；③一些可移动因子或重复序列的序列残留物参与基因的表达调控；④该基因具有已知的功能。以下是在动物基因组中满足上述 4 个条件的实例。

(1)小鼠伴性蛋白　人的基因组有两个串联的补体 C4 基因，即 C4A、C4B。然而，小鼠具有两个相关的紧密串联的类 C4 基因，但是只有一个基因产生类 C4 蛋白；另外一个基因受到不同的调控，只有雄鼠产生伴性蛋白。这个小鼠基因的雄性依赖是由子反转录病毒相关的因子插入 5′区造成的。在雄激素应答位点控制下，转录的起始发生于长末端重复序列。一些证据表明，在小鼠体内，伴性蛋白具有重要的功能。被插入的可移动因子、反转录病毒或多或少地仍不完整，但两个长末端重复序列已发生突变，且彼此不同。编码区受到较大破坏，因此，可得出这样的结论，插入发生在很久以前，有功能的伴性蛋白基因以及在长末端重复序列中的一定控制区序列被选择性地保存下来。Stavenhagenn 说："这个转座子与伴性蛋白联合调控提供了一个在进化过程在被保存下来的长期探索的插入突变的实例"。这个实例符合上述 4 个条件。

(2)RSR 增强子和海胆 *spec* 基因　海胆的 4 个已知基因在它们的 5′区具有"RSR"重复序列。RSR 序列长度超过 600nt，带有约为 200nt 的正向末端重复序列，以两个 RS 符号表示。*spec2a* 基因的 RSR 元件已得到详细的检验，它包括一个增强子，该增强子被定义为 188nt RSR 增强子。它需要 *spec2a* 基因的特定组织表达，并包括 4 个海胆的 *bicoidclass* 核蛋白的 Otx 结合位点。人们研究发现 RSR 区是这个基因的调控中心，基因组中有许多 RSR 元件的额外拷贝功能的 *Alu* 序列，它们被称为中度重复序列。由于它们彼此之间已有明显的差异，并出现在不同的位置，拷贝数也不同，因此，认为在 4 个 *spec* 基因 5′区内的 RSR 元件在很久以前被插入。RSR 拷贝可能来自一个原始的 *spec* 基因内的 RSR，后来这个区复制多次并参与基因调控。

(3)*Alu* 序列中的 DNA 结合位点帮助调控 CD8 基因的转录　Hamber 等检测核酸酶 I 的超敏位点时，在 CD8 基因的最后内含子中发现一个 *Alu* 序列，CD8 基因即为一个超敏位点，并作为对淋巴细胞特异的增强子一部分。这个 *Alu* 序列即含有核蛋白结合位点也明显地对超敏性负责。经过检测发现，在 *Alu* 序列内有 4 个转录元件结合位点是有效的。其中两个 LgF－1 位点、一个 bHLH 和一个 GATA－3 位点。*Alu* 序列中明确被选择改变的证据显示，这个序列在这个位置已适应其功能。证据如下：将这个 *Alu* 序列与它可能的同源基因比较，发现有 7 个不同，其中是 4 个刚提到的，选择性改变的两个是在 GATA－3 位点。这个 *Alu* 序列与现代一致序列和过去的 II 类一致序列分别有 4%、8% 的差异。这样看来这个序列是在 3 000～5 000 万年以前的 II 类主要插入片段插入发生之后，在灵长目动物血统相对近的进化过程中某一时期被插入的。

(4)具有雌激素依赖受体转录增强子功能的 *Alu* 序列　雌激素应答启动子和增强子是通过酵母选择系统从人的 DNA 文库中挑选出来的。雌激素应答区是在古老的 II 类家族的 *Alu* 序列内，与古老的一致序列和现代的一致序列的差异分别是 22%、27%。通过序列比较，该结合位点与已知的雌激素受体位点一致，该位点以 GGTCA 起始，带有它的回纹结构 TGACC。*BRCA-1* 基因也是雌激素应答性基因。目前，*Alu* 序列被认为是在含有雌激素受体细胞内基因转录调控的重要贡献者。

(5)*Alu* 序列是 *WT1* 基因的阻抑蛋白(repressor)　Wilms 瘤 1(*WT1*)是在限定的系列组织中正常表达的基因，但是经常在各种瘤内高效表达。距启动子 12 kb，*WT1* 基因的第三个内含子含有一 *Alu* 序列，这个序列已被证实是一个能降低基因表达的遏止子。这个遏止子在许多被检测的细胞系中以不依赖位置和方向而发挥作用，但在肾细胞内失活，该 *Alu* 序列与 II 类和现代一致序列的差异分别是 17%、20%。现已研究发现在 *Alu* 序列内的 15 个短的 DNA 序列可能充当已知转录元件的结合位点。*WT1* 基因的功能尚未确定，它能在小鼠的特定组织、于特定时期表达。

通过以上分析可以看出，*Alu* 序列是突变和变异的主要来源，*Alu* 序列的不同家族结合不同类的核蛋白。在它们存在的过去的 5 千万年至 1 亿年之间，*Alu* 序列已为许多基因发挥功能作出了贡献，以及作为突变和变异的根源，它们强烈地影响哺乳动物的进化。必须强调的是象 *Alu* 序列这样的重复序列是插入基因区的许多类型的元件之一，而且能影响转录的调控。至今，对插入到动物基因组的元件的认识表明，5 种不同类的插入元件已具有持久的作用，且还有更多的插入元件将可能得到鉴定。植物的证据表明一些其他元件在进化历史中具有较大的价值，DNA 序列插入是否是基因表达中进化变异原始的根源，还有待于进一步研究。

第三节　基因的分子进化[25-31]

基因的进化机制主要是通过基因扩增和不均等交换形成多拷贝基因，再通过突变积累、基因重排和自然选择等因素形成多成员的基因家族或形成新的基因。这些过程可以反复地进行。

一、基因重复对基因进化的影响

基因重复是产生功能分离的一种机制，这种机制常常用于研究哺乳动物基因的进化上。重复或复制转座都涉及到一段 DNA 序列在转座前的一次重复。哺乳动物基因组的 DNA 转座常常通过一个 RNA 中间物发生，并常常导致大量的散布重复家系。被 RNA 聚合酶 II 转录的基因中间复制物，常常缺少功能调节序列，但也有很少是例外的。被 RNA 聚合酶 II 转录的基因通常在转录的序列中含一个启动子序列。这样，这类基因的 cDNA 复制物能达到一个很高的拷贝数，这是因为内部的启动子序列能使转录在复制物上继续进行。这也是 *Alu* 重复家族的进化方式。串联重复基因的产生常常是由于不等交换或不等姐妹染色体交换所导致的。许多簇的人类基因家族就是通过这种机制产生的。但也有一些重复基因简并到非过程性假基因中去，但是，在功能性复制基因和非功能

性假基因之间的过渡可能是一个渐进的过程。假基因是指一个基因在 mRNA 水平或甚至在多肽水平上被表达，但仍是非功能性的。功能的缺乏意味着保存功能的选择压力将降低并最终突变积累导致基因表达的停止或导致不同的表达方式。例如，一些分散的重复基因可在不同的环境中表达。在 α-珠蛋白的基因簇和 β-珠蛋白的基因簇中，分散在不同基因中的序列可能导致产生生物学特性不同的编码产物。编码不同组织同工酶（同一蛋白质的不同形式）的基因也是通过基因重复而形成的。例如，碱性磷酸酶至少被 4 种不同的基因所编码，这些基因在表达上显示了组织特异性。在这 4 种基因中，有 3 种分布在端粒 2q 附近：ALPI 和 ALPP 分别是在肠道和胎盘中发现的编码酶的不同形式（有 87% 的蛋白质序列相似）；ALPPL 编码一种类似胎盘中的酶。第四种酶是 ALPL 位于端粒 1p 附近，编码肝、肾、骨等组织的酶，这种酶和肠道、胎盘中发现的酶关系较远（分别是 57%、52% 的序列相似）。由于基因重复可能是由于原始染色体组的重复所致，编码同种蛋白质不同亚细胞结构的重复基因明显地位于不同的染色体上。例如，在人类肝脏中有两种乙醛脱氢酶的同工酶，一种位于细胞质，另一种位于线粒体。在 500 个氨基酸片段中有 68% 的序列相似。细胞质型和线粒体型分别由位于染色体 9q 的 *ALDH1* 基因和 12q *ALDH2* 基因所编码。

DNA 序列家族基因组内（组间）的交换导致一致进化：在特定基因和 DNA 序列中，同一种内个体成员（同生同源）之间的关系要比不同种直向同源个体间的关系更近一些。因此，如果我们考虑两个种 A 和 B 之间的特殊基因家族，一致进化意味着在 A 品种的一个家族成员与品种中别的家族成员的关系要比 B 品种中相同家族的任何别的成员的关系更密切。一致进化的产生是由于同一染色体组内非等位基因的 DNA 序列的交换而引起的。其中，包括不等交换，不等姐妹染色体交换等。这在串联重复 DNA 序列中特别普遍，例如，不等交换和不等染色体交换能导致一个特殊的重复序列，通过一系列串联重复而扩散，并最终代替别的重复而导致序列匀浆。一些基因家族没有表现明显的一致进化，这是由于在直系同源之间比同一种族内不同家族成员之间的序列同源性更强。通常，位于同一基因簇的基因家族或超家族成员比位于不同基因簇的成员有更高的序列同源性，而且，在同一个簇内相邻基因之间的序列同源性更大。因此，导致任何一个簇形成基因重复常常是近代的串联重复，而已产生不同簇的基因重复则常常是相对远古或从古代基因重复而产生的。

伴随着基因重复，在重复基因之间常存在两种竞争力量影响序列的同一性：序列分散超异（重复基因序列可能在开始是完全相同的，但在进化过程中由于突变在两基因中的积累而逐渐形成差异）和序列均一性（sequence homogenization，在两个基因之间的序列交换可能导致它们的序列混合，因此，维持了序列同一性）。这种均一性是由基因机制导致的（如不等交换，不等染色体交换和基因转变）这种现象在串联重复基因中要比相距较远或非同源重复中普遍得多。因此，相距较远的基因或非同源基因的序列比串联重复基因的序列更容易离散。珠蛋白基因的超家族提供了一个例证，来自不同簇的基因和基因产物之间的同源序列远少于来自同一簇的基因和基因产物之间的同源序列，这很大程度上是由于不同簇被认为是在进化早期已经形成的，而簇内的基因重复则发生在较近的时期。例如，一些重复被认为是不久前产生的，则重复基因几乎是完全相同的。两个人类 α-珠蛋白基因 *HBA1* 和 *HBA2* 编码完全相同的序列，而 γ-珠蛋白基因 *HBG1* 和 *HBG2* 编码

的产物有一个氨基酸不同。

与串联重复基因相比，珠蛋白基因簇内序列交换是不常见的（很近的重复基因除外，如 *HBA1* 和 *HBA2*）。这是因为不同的珠蛋白基因是很小的（1.6 kb），另外，染色体上DNA将它们分离以至不能很好地保存，从而减少在同一簇内不同型基因之间的序列交换。这样，在远缘物种的直系同源成员间如老鼠和人类的同源序列可能多于同一人类簇中的基因间的同源序列，例如，在人和兔的β-珠蛋白的基因间甚至在人和鼠的β-珠蛋白基因间的同源序列要比人类的β-珠蛋白和ε-珠蛋白基因间的同源序列多。

二、外显子重排对基因进化的影响

早在1978年，W. Gilbert 曾指出，通过哺乳动物基因组的外显子重排（exon shuffling），能够产生新的基因，从而产生新的基因产物，这种产物具有新的功能，从而使生物体能更加适应复杂的自然环境。而内含子就是这种外显子重排的历史痕迹，而且内含子的存在可能以某种方式促进了这种外显子重排过程。今天，大量的证据表明，在脊椎动物出现之前的5亿年中，外显子重排在基因构建过程中起着主要的作用。

关于内含子如何起源的问题，目前有两种假说，以哈佛大学 W. Gilbert 为代表的一派认为，内含子与它所在的基因一样古老，在装配第一个这样的基因时，内含子就已存在。而今天的原核生物和少数低等真核生物，由于它们需要进行快速的DNA复制而进行快速的细胞分裂，因而失去了内含子序列。而对于绝大多数真核生物来说，它们没有必要进行快速的DNA复制和细胞分裂，因而将内含子序列保留下来，它对于保持基因的相对稳定又有重要作用。这种观点认为，今天高等真核生物基因中的内含子更具有代表性。原核生物中的内含子的发现有力地支持了这种假说。

以变化缓慢的5S rRNA基因为标准，真核生物大约在15亿年前从原始细胞中分化出来；12亿年前真菌就与动植物祖先分家；10亿年前，动植物分家；4亿年前出现鱼类；3亿年前出现了哺乳类和鸟类。

以 F. Crick 为代表的一派认为，不连续基因的内含子不是该基因本来就有的，而是在进化的某一过程通过转座作用插入到连续基因中去了。前面所提到的酵母线粒体基因的内含子编码的转座酶所产生的基因转化，有利地支持了这种假说。1987年，Littman 等人发现人和鼠的T细胞表面糖蛋白 CD4 分子中，有一段免疫球蛋白分子中的类V结构域（V-like domain）；而在 *CD4* 基因的类 *V* 结构域的编码序列中存在一个内含子，可是在含有类 *V* 结构域的其他众多成员中至今尚未发现有这样的内含子序列。这很有可能是在 *CD4* 基因进化中，在V结构域插入此内含子。

三、不均等交换对基因进化的影响

我们已经看到，各种哺乳动物中的珠蛋白基因都成簇排列，但在活跃基因和假基因的数目上有很大差异。这主要是由不均等交换而产生重复基因，其机制与重复序列的产生相同。

在不连续基因中，由于外显子受到选择的压力，因而突变概率很小，而内含子不受这

种压力，因而突变概率大。这样，交换往往在外显子部位发生。由于内含子在序列组成和大小方面分化很快，从而限制了这种不均等交换的发生，保持了基因簇的相对稳定。这可能是内含子存在的生物学意义之一。然而不均等交换是不可避免的，从人和鸡的珠蛋白基因簇的人工克隆中经常可以分离到缺失的基因簇。由不均等交换产生的缺失和重复对机体往往是有害的，就被自然淘汰；但也有个别的不均等交换产生了更大的选择优势，产生了新的活跃基因，原有的基因丢失，或保留，有的因突变而变成假基因。哺乳动物现在的珠蛋白基因簇状况就是这种不均等交换、突变、重排和自然选择的结果。

根据各物种的珠蛋白基因组分，可以推测珠蛋白祖先基因的可能进化途径。见图 14-5。

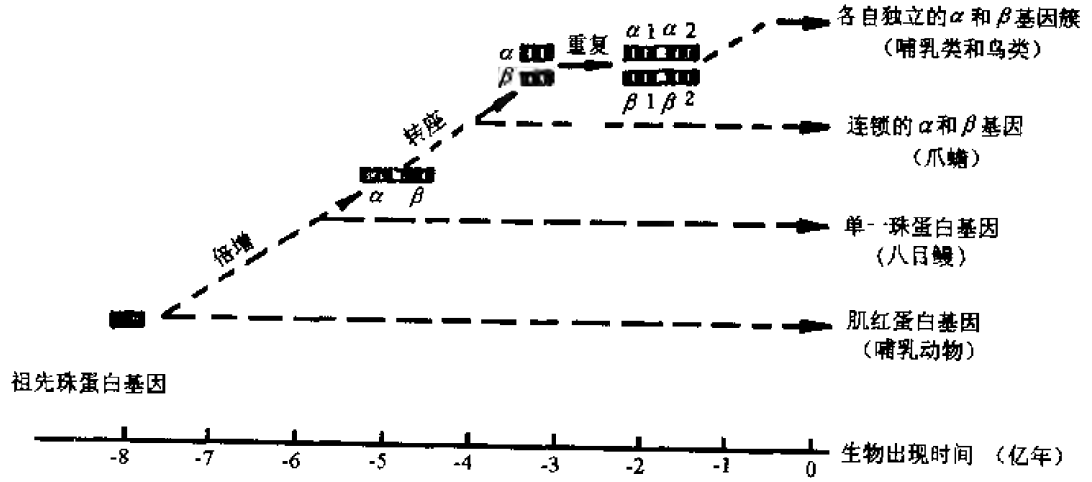


图 14-5 珠蛋白基因由一个祖先基因通过扩增转座、重复而进化的可能途径

哺乳类和鸟类都具有各自独立的 α 基因簇和 β 基因簇，而两栖类的 α 基因和 β 基因仍然是连锁的。哺乳类和鸟类从爬行类祖先歧化出来的时间大约在 2 亿 7 千万年前，而爬行类从两栖类祖先歧化出来的时间大约在 3 亿 5 千万年前。这样，从 3 亿 5 千万年到 2 亿 7 千万年前的时间可以定为 α 基因和 β 基因失去连锁的时间。α 基因和 β 基因分离可能是由转座作用造成的。某些原始鱼类只有一种珠蛋白基因，这些鱼类一定是在珠蛋白基因倍增并分化成 α 基因和 β 基因之前从进化路线上歧化出来，这个时间大约在 5 亿年前。也就是说，珠蛋白基因的倍增和分离是在 5 亿年前至 3 亿 5 千万年前之间。从氨基酸组成序列的差别可以推算出珠蛋白基因和肌红蛋白基因分化的时间在 8 亿年前。豆科血红蛋白基因和珠蛋白基因分化的时间更早。

把不同的生物中某种蛋白质的氨基酸组成序列进行比较，发现两种生物之间差异的百分数大约和它们自同一祖先向不同方向演化的时间成正比，也就是说，突变在生物进化过程中发生和积累的速度大体上是恒定的，这种情况常常被称为进化钟（evolutionary clock）通过比较不同生物种类的珠蛋白基因的变化情况，就可以得出珠蛋白基因的进化速度大约为每 1040 万年有 1% 的碱基出现替代型差异。用这一速度就可以推算珠蛋白基因在种内形成变化的大致时间表。例如人类 β 簇中的 β 基因与 δ 基因间有 3.7% 的碱基替代，因此这两个基因分化出来的时间在 3850 万年前。

不同的基因承受着不同的负选择压力，这就决定了生物钟的快慢，因此不可用一种基因的进化速度去推测另一种基因的进化时间表。一般来说，基因的产物越重要，对它的结构和功能要求越严格，变异积累的速度越低。同一个基因的不同区段的重要性不同，所承受的负选择压力也不同。因此外显子的序列变化就慢得多。同是外显子，决定蛋白质活性部位的碱基序列比其他序列要稳定得多。

基因的重复减弱了负选择压力，使相同基因中的一个或几个能较快地积累突变从而形成新的基因家族成员或变成失活的假基因。一旦成为假基因以后就完全不受选择压力的影响，变异积累的速度就大大地高于活跃基因而最终面目全非。假基因的命运或者是因 DNA 缺失而消失，或者是作为一般无用的 DNA 单一序列而存在下去。然而，有些基因在重复后的长期进化中却一直保持着几乎完全一致，如人类珠蛋白的两个 α 基因和两个 γ 基因，现在普遍认为，这种非等位的相同基因在遗传过程中是作为一个单位而传递的，它称为重合进化(coincidental evolution)，有时也称为共进化。解释一致进化现象的两种机制即交换固定学说和基因转换学说，不管哪一种机制，其实际结果是：突变的基因要么被淘汰，要么占据全部的原有相同基因位置。

同一基因家族的不同成员之间的碱基差别是由突变引起的。这些突变的频率很低，产生的后果大多有害，很快被负选择压力所淘汰。也有一部分被固定到生物群体中。那么如何解释从个体发生的突变扩展到整个物种中呢？经典的理论是达尔文的适者生存理论，即突变体具有更大的正选择优势，适应力强，繁殖得更多更快。而一种由赖特提出的遗传漂变(genetic drift)学说认为，群体较小时，自然选择不起作用。日本群体遗传学家木村资生提出了类似的中性漂变学说。这种学说认为，从个体到群体的扩展完全是中性突变(无利也无害的突变)。一个中性突变能否扩展到整个种群，基本上或完全取决于机会。事实上在 DNA 水平上的许多突变都是中性的，包括不改变氨基酸组成的同义突变以及不影响蛋白质空间结构和功能的无声突变。这一学说得到了实验群体遗传学的支持。但是否能反映生物进化的实际情况，还有待于研究。

第四节　基因组的起源和进化[32－39]

对基因组的起源和进化的研究，是探索生命起源和进化的重要组成部分。根据现代基因组在细胞中的地位和功能，可以把它们分为自主性的核基因组、处于细胞器中半自主性的线粒体基因组和叶绿体基因组。从进化的角度看，这三种基因组都是源于一种原始基因组。因此，对现代各种基因组的结构进行深入分析，找出原始基因组的“遗迹”及其变化趋势，是研究基因组起源和进化的有效途径。

一、核(类核)基因组的结构与进化

1. 核(类核)基因组的结构与进化

核基因组和类核基因组在细胞中的地位是等价的，在进化上必须一起考虑。但从结构方面看，两者相差甚远，可以很自然地把它们分开。原核生物的类核基因组一般是环状，再形成类核高级结构。通常基因组大小只有数百万碱基对。类核基因组都是由单拷

贝或低拷贝的DNA序列组成，基因的排列比较紧密，一般以多顺反子为转录单位，较少非编码序列，非编码的重复序列和内含子就更少了。至今，只在盐杆菌(*Halobacterium*)、硫氧化菌(*Sulfolobus*)和脱硫球菌(*Desulfurococcus*)等原细菌的基因组中，以及在部分真细菌的基因组中发现少量各异的重复序列和内含子。至于真核生物的核基因组，一般是多线状，再形成染色体或间期核的高级结构，大小要比类核基因组大好几个数量级，且变化范围很大，已知最小的多细胞动物核基因组约为4×10^5 kb，最小的高等植物核基因组也有7×10^4 kb。在结构组成上，核基因组以普遍存在大量非编码序列为特点。其中重复序列一般可占基因组的1/3，在有些植物核基因组中甚至可达90%。内含子也占基因组相当重要的比例，真核基因基本上都是有内含子的断裂基因。一个基因的长度可有多至90%的内含子。重复序列和内含子的含量直接影响了基因组的大小。

因此，在大小和构成上，类核基因组是"节俭型"的，而核基因组是"冗余型"的。节俭型与冗余型的差别在于基因组的大小以及重复序列和内含子的含量，所以可把它们分别称为"小基因组"型和"大基因组"型。"小基因组"型以基因组较小，不含或少含重复序列和内含子为特点；"大基因组"型则以基因组巨大，含有大量重复序列和内含子为特点。

关于核(类核)基因组的进化，现已有越来越多的证据表明核(类核)基因组是比较直接地从原始基因组进化而来。而原始基因组是由重复序列组成，并在这样的一个基础上发展成具有原始的外显子、内含子及重复序列的原始基因组。原始基因组再向着小基因组与大基因组两个方向发展。如果向小基因组进化，则由于繁殖效率、空间结构等的约束，不再留有重复序列和内含子结构(它们在进化中丢失)；如果向大基因组进化，则重复序列、内含子作为残遗结构都能继续存在，并起着结构上和进化上的功能。在现代生物中，原核生物的类核是小基因组进化方向的代表，而真核生物的核则是大基因组进化的代表。

2. 染色体组型的进化

染色体组型在生物进化的过程中所发生的数目和结构上的变化，无疑在物种形成中起着重要作用。在植物界的高等植物中，染色体结构上的重新排列和多倍体在染色体组型的进化中发挥了重要作用。而动物中染色体变异的主要源泉看来局限于构造上的重新排列，例如易位，特别是罗伯逊类型的易位。

在动物中，多倍体物种的稀少并不是意料之外的，这因为大部分动物有两性之分，它们有性染色体的作用，这种作用会被多倍体所否定。大多数高等植物是雌雄同体的，它们并没有性染色体。

一个典型的二倍体动物的物种有两套常染色体(2A)和一个XX(雌的)或XY(雄的)性染色体。如果这样的物种变成四倍体，那么雄的个体会是4A,XXYY，而雌的个体会是4A,XXXX。在减数分裂中，雄的XXYY染色体会形成一个XX二价体和一个YY二价体，结果所有雄性的配子会含有XY性染色体的组成，即2A,XY。因为雌的只产生出2A,XX的配子，那么由交配所产生四倍体会含有性别难于决定的染色体组型：4A,XXXY。这样，在两性动物中的性染色体机制大概就成为多倍体在动物界中进化的主要障碍之一。虽然如此，多倍体可在雌雄同体的低等动物中见到，它们由孤雌生殖方式产生后代。

有人指出，多倍体曾发生于哺乳类早期的进化中。关于这方面并没有确实的证据。但是在某些哺乳类的物种中包括人类在内，有些非同源染色体的形态相似及染色体带型相似支持了上述观点。其中一个例子就是人的 4 号染色体和 5 号染色体之间的极其相似。

正如前述，在动物物种中，染色体组型进化的最显著的源泉是通过倒位或易位所发生的染色体结构的变化。果蝇中密切相近的物种之间染色体变异最常见的形式是臂内倒位，虽然有时也看到着丝位的融合和臂间倒位。在果蝇属中，约有近 2000 个物种具有达 7 万个臂内倒位类型的重新排列。

哺乳类的物种中，已知有许多类型的染色体结构异常，但是在形成中最有意义的是着丝粒融合型的易位（罗伯逊易位），在那里两个具端着丝粒或具近端着丝粒的染色体的着丝粒合并在一起，形成一个具中间着丝粒的染色体。在具近端着丝粒染色体的情况下，靠近着丝粒的结构异染色质的区域发生了断裂，接着是染色体的两个长臂融合成一条大的具中间着丝粒的染色体，另一个易位的产物就是一个细小的中央片段，由结构异染色质组成，它会丢失，但不会引起表型的变化。这一类型染色体结构变化的结果不仅是染色体组型外表的结构变化，而是二倍体组成中染色体数目的减少。在技术上这可以看成是产生了非整倍体，但是染色体臂的基本数目是保持不变的。

作为不同着丝粒融合的累积结果，在哺乳类非常相近的物种之间的染色体组型，其变异可能是很大。例如美国刚毛棉鼠有一个二倍体的数目是 52，但是有一个密切相近的物种中着丝粒融合把染色体数目减少到 22。

上述所提到的染色体组型进化的例子都是相近物种的染色体组型，它们的染色体数目发生了变化而染色体臂的数目却保持稳定。但是，某些类群的动物中却发生了相反的情况，即物种之间的染色体数目保持不变，而染色体臂的数目却发生了变异。已经知道，这样的变异可能是臂间倒位或是不平均的相互易位的结果。但是最近知道还有另一种可能的机制，即由于获得或失去另外的臂，而这些臂主要是结构异染色质组成的。这后一情况的例子曾见于鹿鼠，其属名是 *Peromyscus*。在所研究的 20 个物种中，二倍体的染色体数目总是 48，但是臂的总数变异在 56～96 之间。应用 G 和 C 带的分析表明在这个物种 *Peromycus crinitus*（$2n=56$）和另一个物种 *P. ceremicus*（$2n=96$）中其常染色质内含基本一致，而所增加的染色体臂则完全由遗传上不活泼的结构异染色质所组成。如此大量的结构异染色质的来源及其在染色体组型进化和物种形成中的作用仍是一个谜。

二、线粒体基因组的结构与进化

1. 线粒体基因组的结构

作为一个能自我更新的细胞器，线粒体含有独特的基因组。在其组成上，一般有编码线粒体内大小亚基 rRNA 的基因、全部或部分线粒体 tRNA 基因、部分呼吸链特有酶的基因。线粒体的功能在各类真核生物中都相同，但其基因组却变化很大。多细胞动物的线粒体基因组十分细小和致密，绝大部分的大小都在 15.7～19.5 kb 这个范围内。基因组呈环状，没有或很少间隔序列，基因中没有内含子。不存在非编码的重复序列。

相反高等植物线粒体基因组是最大、最复杂的，基因组呈单环或多环状，大小范围从

186～2400 kb。由于植物线粒体基因组所含的基因只比动物的多两倍左右，因此，有很多基因间隔，另外还有很多重复序列和内含子。

原生动物线粒体基因组的大小在 18.5～55 kb 之间，呈环状、线状、或由大环与小环组成的网状结构。动基体(kinetoplastida)种类的动基体大环 DNA 除了有一段 1517 kb 的转录区外，还有由长度不等的重复序列组成的多变区域，其他一些原生动物线粒体基因组中也有少量重复序列，不过还没有在研究过的基因组中发现内含子。

对藻类的线粒体基因组研究不多，已知莱因衣藻(*Clamydomonas reinhardtii*)的 158 kb 基因组的结构像多细胞的而不像高等植物的。

表 14-2 真核生物线粒体基因组的大小和构型

物种	大小(kb)	构型*	物种	大小(kb)	构型[1)]
脊椎动物、昆虫			原生动物		
Homo sapiens	16.569	环状	*Plasmodium lophurae*	18.5	环状
Mus musculus	16.259	环状	*Tetrahymena pyrijormis*	55	线状
Gallus domesticus	16.2	环状	*Trypanosoma brucei*	20～22	大环
Drosophila neohydei	15.7	环状	藻类		
D. melanogaster	18.7～19.5	环状	*Chlamydomonas reinhardtii*	15.8	线状
高等植物			*Euglena gracilis*	60	—
Marchantia polymorpha	186.608	环状	*Olisthodiscus luteus*	40	环状
Brassica hirta	208	环状	真菌		
B. campestris	218	三环	*Aspergillus nidulans*	32～33	环状
Pisum sativum	370	—	*Cochliobolus heterostrophus*	115	环状
Cucumis sativus	1500	—	*Hansenula mrakii*	55	线状
C. melo	2400	—	*Sacchuromyces cerevisiae*	68～77.8	环状
Zea mays	570	多环	*Schizosaccharomtyces pombe*	17.6～23	环状

* 近年来发现高等植物和真菌的线粒体基因构型在大多数情况下是线状的。

真菌的线粒体基因组大小在 17.6～115 kb 之间，绝大部分呈环状。基因组的大小主要由基因间隔的长度、内含子及重复序列等决定。随着基因组变小，这几种非编码序列，特别是内含子也减少。但即使在最小的基因组中，也有个别基因具有内含子。

概括起来，多细胞动物和高等植物线粒体基因组的差异与类核基因组和核基因组的显著差别类似。它们代表了线粒体基因组在大小、结构上的两种形式。因此也可以把它们称为(线粒体)"小基因组"型(动物型)和(线粒体)"大基因组"型(植物型)。"小基因组"型以线粒体基因组细小、致密，基本上不含非编码序列(重复序列、内含子)为特点；"大基因组"型则以线粒体基因组觉得较大，含有很多非编码序列为特点。至于原生动物、藻类和真菌的线粒体基因组，一般来说，基因组较小的偏向子动物型；基因组较大的则偏向于植物型。

2. 线粒体基因组的进化

线粒体基因组也是从原始基因组进化而来的，但从原始基因组进化成线粒体基因组需要经历一个内共生过程。已进入内共生状态的线粒体祖先的基因组在进化过程中，一方面把相当部分的基因转移到核中(同时也接受一些来自核、其他细胞器的 DNA 片段)，另一方面就沿着小基因组方向与大基因组方向两种途径发展。

进化途径是基于进化可能性的。在进化过程中，对生命始终存在着效率方面的选择压力。如果没有其他限制，生物对这种选择压力就可有两种应答方式：或者变得有更高效率、更快繁殖，使这种选择压力得以加强，从而使向高效率的进化成为一种趋向；或者变得更复杂、更具多功能，从而减少了效率方面的选择压力，使进化可以继续向复杂性和多功能方面发展。这样，基因组大小的进化可能性就成为两种进化途径：或者向高效的"小基因组"发展，或者向复杂的"大基因组"发展。如果效率选择压力对不同的生物原本就有差异，则同样会造成相似的进化结果。因此，原始基因组若向"小基因组"方向进化，由于效率选择压力大，而且有空间大小的约束，这种进化使基因组趋向于只含必需序列，非必需的重复序列和内含子在进化过程中就被丢掉了。若向"大基因组"方向进化，由于效率选择压力小，这种进化使基因组中的重复序列和内含子可以与必需基因共存及发展，并起着结构上与进化上的作用，同时也使基因组大小的变化范围可以很大，这两种进化趋向正是各种生物基因组共有的进化途径，即分别是"小基因组"和"大基因组"途径。"小基因组"进化途径以丢失、残留重复序列和内含子为特点，包括类核基因组、动物型线粒体基因组以及大部分藻类和高等植物的叶绿体基因组的进化；"大基因组"进化途径以保留、发展重复序列和内含子为特点，包括核基因组、植物型线粒体基因组的进化。这样，通过这两种进化途径就把现代三种基因组的结构特点的产生予以说明，并把它们的进化统一起来了（表 14-3）。

表 14-3　3 种基因组的结构共同点及共有的两种进化途径

基因组进化	"小基因组"途径	"大基因组"途径
核（类核）基因组线粒体基因组	原核生物后生动物、具较小线粒体基因组的低等真核生物	真核生物高等植物、具较大线粒体基因组的低等真核生物
叶绿体基因组	大部分藻类和高等植物	以伞藻属为代表
原始（核、线粒体、叶绿体）基因组 ↓ 现代（核、线粒体、叶绿体）基因组	在同一种基因组中属小型，基因排列紧密　重复序列和内含子在进化过程中丢失，因此，不含或较少含这些成分	在同一基因组中属大型，基因排列分散。在进化过程中保留并发展重复序列和内含子

既然原始的核（类核）基因组、线粒体基因组和叶绿体基因组都有内含子，那么在这三种基因组的"小基因组"进化途径中，应该都存在着去除这些成分的机制。实际上，内含子在某一基因内存在与否具随意性，DNA 剪接可通过整合酶催化的逆反应而实现。这些事实说明这种机制在进化中是存在的，它包括剪接、反转录、基因转换等。

三、展　　望

通过对生物基因组结构的分析，可以看出从原始基因组向大、小基因组方向进化这一模式具有一定的普遍性，它不但对核（类核）、基因组的起源和进化能给以较好的说明，而且也能解释现代各类线粒体基因组的结构差异及其起源和进化。重复序列、内含子、核（类核）、基因组、细胞器基因组的起源和进化可以通过一个总的进化模式说明。生物基因组在进化过程中，处于一种动态、革新、发展的变化状态，小基因组型的常常把原始的成分"丢失"了，相反，大基因组型由于在空间上所受的限制或选择压力较小，原始的成分常常

能在一定程度上保留下来。因此，一般来说，对于同一基因组，如果越大，其就可能保存更多原始基因组的遗迹。认识这一点，对于研究基因组的起源和进化是很有意义的（见图14－6）。

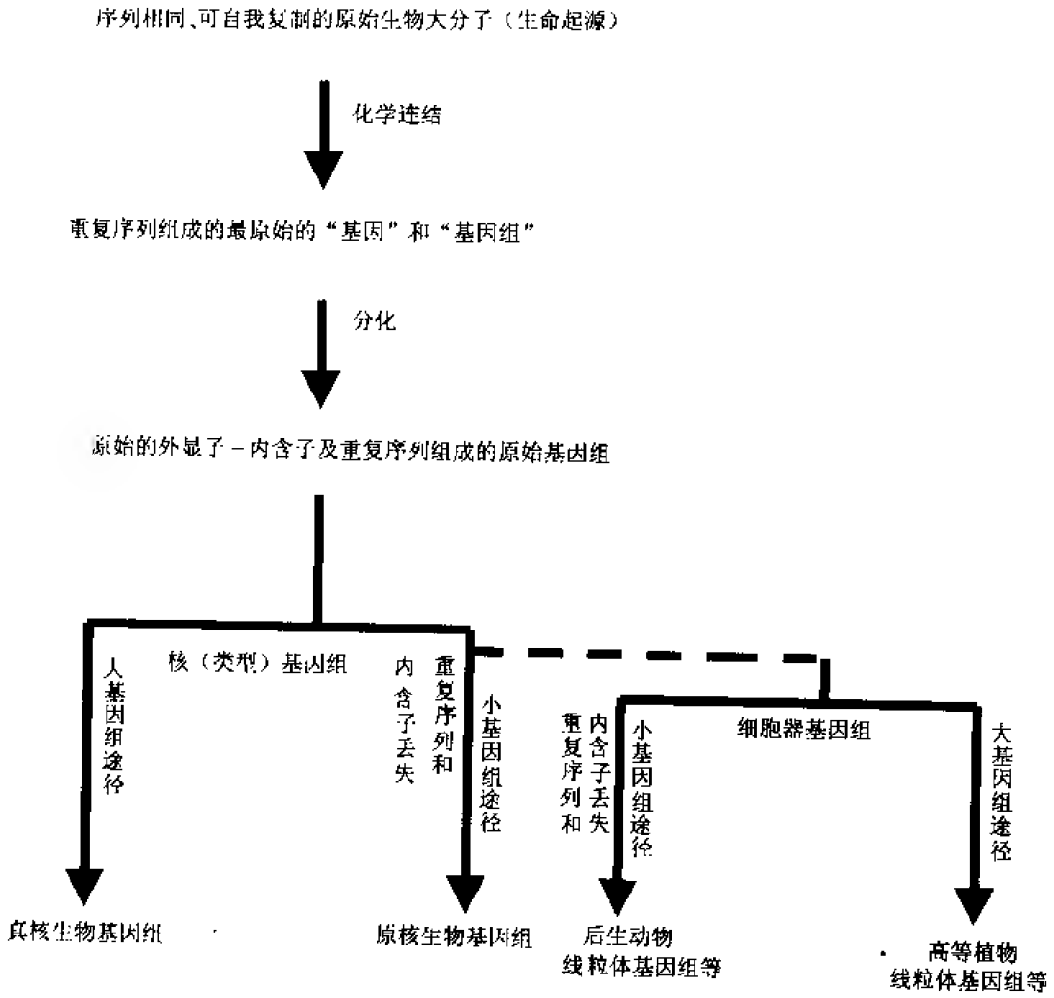


图 14－6　重复序列、基因结构、核（类核）基因组和细胞器基因组的起源和进化

第五节　分子进化[40~44]

一、血红蛋白的分子进化

1963 年著名的生化学家 Ingram 就开始对血红蛋白的氨基酸排列序列进行了探讨。随后的研究大大地加深了人们对这类球蛋白分子进化的认识。

大部分脊椎动物包括人在内，它们的血红蛋白是由 4 条多肽链组成的，是一种四聚体（tetramer）。这 4 条多肽链中有两条是内容一样的 α 链，含有 141 个氨基酸；有两条也是内容一样的 β 链，含有 146 个氨基酸。这种血红蛋白大量地存在于出世后的人体中，经常

写作：$\alpha_2\beta_2$

每条多肽链的内部各含有血红素的分子，由它来运送或贮存氧气。血红素是含铁的卟啉化合物。

人类从胚胎开始出现血红蛋白起到成人，所含有的血红蛋白不止一种，而是至少有以下的类型：

Hb-A(成体血红蛋白)$\alpha_2\beta_2$

Hb-F(胎儿血红蛋白)$\alpha_2\gamma_2$

Hb-A2(成它中少量的血红蛋白)$\alpha_2\delta_2$

Hb-E(胚胎血红蛋白)$\alpha_2\varepsilon_2$

胚胎血红蛋白只见于胚胎期，其后逐渐由胎儿血红蛋白所取代。胎儿血红蛋白在婴儿出生后就逐渐消失，被 Hb-A 所代替。$Hb\text{-}A_2$ 只占成人血红蛋白中的少量。

在一些植物、无脊椎动物及无颌类脊椎动物中，运送和贮存氧气的球蛋白只有一种。这就是肌红蛋白(myoglobin)。在绝大部分脊椎动物中，则有两类球蛋白，即肌红蛋白和血红蛋白。肌红蛋白存在于肌肉中，使肌肉呈暗红色。血红蛋白则存在于血液中。

据分析，肌红蛋白是单链结构。这一条链内的氨基酸种类及排列序列和 α 链有很多相似之处。这暗示着这两条多肽链有同源性。现在，已知它们分别由不同的基因所控制，这暗示着这两个基因(它们是不同染色体上的不同位点)同源，即它们有共同的祖先。

β 链与 α 链的内容很相似。这暗示着它们有共同的起源。不言而喻，有关的不同基因之间也有共同的祖先。

γ 链、δ 链和 ε 链的内容与 β 链很相似，例如人的 β 链与 γ 链有 39 个氨基酸的差异，而 β 链与 δ 链只有 10 个氨基酸的差异。

已经知道，α 链、β 链、δ 链、γ 链等都各有自己的基因。显然，这些基因彼此有较相近的亲缘关系。

那么，肌红蛋白和血红蛋白的基因是如何起源的呢？

现在认为，基因起源的一个重要途径就是通过有关基因或染色体片段的重复和分化。

已经知道，重复和缺少是染色体突变相互联系的两种类型。在减数分裂中，同源染色体在前期中要配对，随后彼此分开。在分开过程中，就经常发生互换。在互换中，就可能出现某些差错，出现了重复和缺少；即一条染色体多了一段 DNA(或多了一个或几个基因)，另一条染色体则少了一段 DNA(或失去一个或几个基因)，以后，经过另一世代或若干世代的生殖作用，重复基因就可以成为：AA。这些基因有同样的功能。

其中一个 A，如果发生突变，就可能改变原来突变基因的功能，使基因 $A \rightarrow B$。于是成为：AB。这就是基因通过突变而发生分化。以后通过易位，可以使基因 A 和基因 B 分别分布在不同的染色体上。这样，基因的多样性就实现了。由一个基因就产生出若干个基因了。肌红蛋白基因和血红蛋白基因大抵就是这样来的。

这可能发生在什么时代呢？

运送氧气的球蛋白来自某个原始基因。这个基因大抵出现在 6 亿年前。那时多细胞动物已经出现了。这是现在所知道的肌红蛋白基因的祖先。

大约在 4 亿年前或更早些，由于基因的重复和以后发生的突变，就出现了原始的血红蛋白基因。这大概是链的基因。原始的血红蛋白可能像肌红蛋白那样是单链的，以后才

出现了双链结构，如 α_2，以后才有四聚体如 α_4。这改进了血红蛋白的功能。

到了两栖类，血红蛋白都成为 $\alpha_2\beta_2$ 的类型了。

到了哺乳类，才出现了胎儿血红蛋白 Hb-F，即 $\alpha_2\gamma_2$。这是新的事物。

到了猿类和人类，才出现了 $\alpha_2\delta_2$。

二、细胞色素 c 的进化

细胞色素 c(cytochrome c)是一种有关呼吸作用的蛋白质。它一般含有 104 个氨基酸。它存在于有氧呼吸的原核生物和真核生物的细胞质中。它最早可能出现在某些真核生物里，距今约 13 亿年前或更早些。也可能在某些好氧的原核生物里就有它了。

在呼吸作用中，它与血红蛋白有不同的功能。前者在于传递电子，后者在于运送和贮存氧气。它们共同协作，推动有氧呼吸的正常进行。

细胞色素 c 在所有有氧呼吸的生物中，都是与细胞色素酶相互作用，而进行工作的。在不同物种的生物中，它们的细胞色素 c 的氨基酸大抵有一半是相同的。这暗示着这些生物的细胞色素 c 彼此同源。它们有共同的起源。在进化中，细胞色素 c 中的许多氨基酸出现了差异，这显然是由基因突变所引起的。但是这些突变并未影响细胞色素 c 的正常功能。

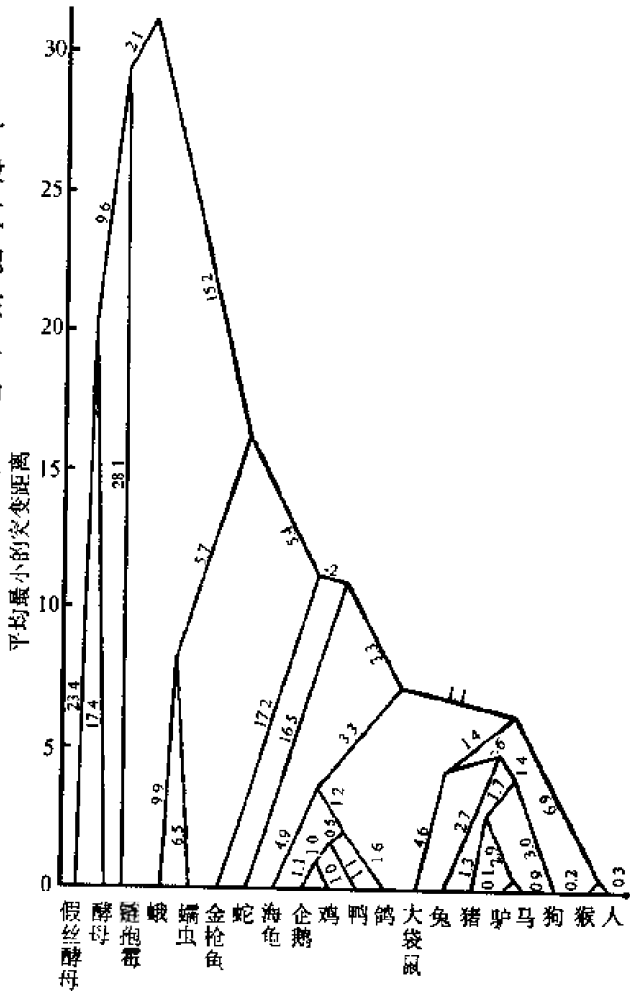


图 14-7　根据不同物种细胞色素 c 氨基酸序列的差异描绘的系统树

例如，已经确定了约 30 种不同物种的细胞色素 c 的氨基酸序列。在每种蛋白质 100 个左右的氨基酸序列中，约有 30 个位置似乎是不变的，其余则有程度不一的氨基酸取代。可以假定，这些氨基酸的差异是由进化过程中单碱基突变引起的，通过每一对蛋白质之间的氨基酸差异的比较，有可能估计出从其最近的共同祖先分化时所需的最低突变数。因而，可根据这些资料描绘出最可信的系统树(图 14-7)。这样得到的系统树，与传统的利用形态学和其他分类标准得到的系统关系非常接近。由于在这种情况下，系统树完全是从一条多肽链的氨基酸序列推断出来的，故可用来表示一个基因的进化。

图上的数字是一种细胞色素的基因变为另一种细胞色素基因时，所需改变的核苷酸

的最小估计值。某些不同生物的细胞色素 c，其所含氨基酸差异的数量如表 14－4 所示。

表 14－4　不同生物与人的细胞色素 c 成分差异的比较

物　种	与人的不同的氨基酸数
黑猩猩	0
猕　猴	1
家　兔	9
袋　鼠	10
鲸	10
猪、牛、羊	10
狗	11
驴	11
马	12
鸡	13
响尾蛇	14
龟	15
金枪鱼	21
鲨　鱼	23
家　蝇	25
蛾	31
小　麦	35
红色链孢霉	43
酵母菌	44

从表 14－4 可以知道，人和黑猩猩的细胞色素 c 所含的氨基酸完全相同。而人和猕猴的细胞色素 c 只有一个氨基酸的差异。随着亲缘关系的距离，氨基酸差异的数量大抵表现了一定的相关。例如，人和家兔的细胞色素 c 有 9 个氨基酸的差异，人和龟有 15 个氨基酸的差异，人和小麦有 35 个氨基酸的差异，人和酵母菌有 44 个氨基酸的差异。

上述的相同和相异，不能归于偶然。有理由认为这是同一基因突变的产物。

需要指出的是：上述的基因突变所引起的细胞色素 c 中氨基酸的差异，并不影响（或不显著影响）细胞色素 c 的功能。这些突变大抵属于中性突变。这是中性进化理论的根据之一。

三、催乳激素的进化

催乳激素（prolactin）是另一种蛋白质，由脑垂体的细胞所产生，也是基因的产物。

这种激素在不同的脊椎动物中，在成分上有一些差异，在功能上则差异很大。例如，催乳激素在哺乳类中，只在雌性个体（母体）产生婴儿时才产生出来。它能促进乳汁的分泌。在其他脊椎动物中也有类似的蛋白质，但是功能彼此很不相同。例如，在某些鸟类中，这种激素能促使鸟去抱卵，使卵能在水里得到受精。这种激素的多样性是怎样来的呢？一个合理的解释是有关的基因在脊椎动物的进化中，发生了若干突变，成为不同基因，并改变了功能，以适应不同的需要。

四、核糖体 rRNA 的分子进化

核糖体是由 rRNA 和蛋白质所组成的。不同生物中的 rRNA 成分是由 rDNA 决定

的。分析不同生物的 rRNA 成分，就可以推测染色体中 rDNA 基因的变化情况。

RNA 的成分可以从分析所含碱基(A、C、G、U)的比例知道。由于 A 与 U 互补，它们的量是一致的。C 与 G 互补，它们的量也是一致的。从 G+C/A+U 的比例变化情况，可以推知不同生物之间的亲缘关系。

分析 6 种单细胞生物的 GC/AU 的比例，彼此各不相同。其中眼虫的 GC/AU 比例与海绵的 GC/AU 比较相近。这与一般的进化观点一致。一般认为，多细胞动物是从眼虫这类原生动物进化而来的，而海绵是多细胞动物中最低级的类型。

值得注意的是还有 GC/AU 的比例分两支变化：从低级软体动物出发为一支，向高级软体动物方向发展；另一支则从棘皮动物出发，向脊椎动物方向发展。

GC/AU 的比例朝两个不同方向发展，这不是偶然的。现代动物学一般认为棘皮动物和脊椎动物有比较密切的亲缘关系，它们有共同的祖先。这个原始祖先在进化中很早就与软体动物分道扬镳地进化了。这暗示着动物的进化也反映在 rRNA 的分子上。这看来是一个合理的解释。

分析细菌、单胞藻、真菌和高等植物的 rRNA 中的 GC/AU 比例，也可以看到相似的情况。这些生物的亲缘关系也记载在它们的 rRNA 分子之中。

这些材料暗示 rDNA 基因的进化。

第六节　群体遗传与进化[45~46]

自达尔文的《物种起源》出版以来，物种的形成与进化研究一直成为生物学界关注的中心问题。美国遗传学家 Dobzhansky 于 1941 年发表的《遗传学与物种起源》，美国生物学家 Mayer 于 1942 年发表的《系统学和物种起源》，英国 Simpson 于 1944 年发表的《进化速度、方式和进化的意义》，美国植物学家 Stebbins 于 1957 年发表的《植物的变异和进化》等等是 60 年代以前生物系统学和细胞遗传学对物种问题研究的总结。他们的研究已经证明动植物分类学家的发现与遗传学家有关进化机制的学说是一致的。依据这一时期系统分类学、细胞遗传学和生态学的研究综合，生物学家提出了一套物种形成和进化的综合理论：即种群是生物进化的基本单位，进化机制的研究是属群体遗传学的范围。突变、选择、隔离是物种形成及生物进化过程中的三个基本环节。这一理论对半个多世纪以来的生物学理论和实践产生了重大的影响，特别是对物种形成的若干理论的一系列新发现，对创造新类型生物的技术和方法上的突破，使生物学在各个领域都取得了划时代的研究进展。

有生命物质最明显的单位是个体，但在进化中重要的单位不是个体而是群体。群体是许多个体的集合体，它们是由交配和亲子系连续起来的。换句话说，群体是同一物种若干个体的共同体。群体的遗传组成可以代代改变 ——发展，它们的不断改变体现着生物进化的可能，因为群体是一个变异库(基因库)，把群体(物种)作为一个整体单位，看成是一个进化着的“个体”，所以生物进化意味着有机种群的多样性和适应性不变化，这种变化是逐渐的、连续的、通常带有定向的成分。群体遗传学的发展为进化的深入研究提供了理论依据和定时分析的手段。

一、进 化 理 论

1. 拉马克的进化观点

比较完整的进化学说最早是在1809年，由法国博物学家和哲学家拉马克(Jean Baptista Lamarck 1744～1829)提出来的。他认为生物的种(species)不是恒定不变的类群，而是由以前存在的种衍生而来的，他首次提出了物种可变的观点。他集中研究了生物长时间的变化过程，为了解释这种进化的过程，他提出：有机体内具有趋向于完善的能力；有变得适应于环境的能力和获得性遗传能力等。他认为环境的改变是物种变化的原因。最明显的例子就是长颈鹿发展进化，最早短头颈的祖先在食物缺乏的环境里，必须伸长头颈来吃高树上的树叶子，因此头颈长得要长一些，这个变化传给了后代。由于环境里仍然食物贫乏，同样需要伸长勃颈才能吃到高树上的叶子，又使子代的头颈长得长一些，这样一代一代下去，长头颈的特性继续加强并遗传下来，终于成为现代的长颈鹿。

拉马克的进化学说虽然推翻了物种不变论，建立了生物学的历史观点，他的学说看来也很容易说明生物的进化现象，但是得不到科学实验的支持。

使拉马克想到进化的许多事实和观察，也引起了18和19世纪其他博物学家的注意。但是拉马克对进化过程的解释太一般化了，令人难以信服。一切都太容易了，以致不能宣称环境改变创造了进化的改变，但是这个过程究竟是怎样的呢？这种前提如何去进行实验？正是查理士·达尔文(Charles Darwin)提供了答案。

2. 达尔文的进化学说

达尔文(1809～1882)从1831年开始，以博物学家的身份参加了"贝格尔"号5年环球航行，他观察到生物学的多种现象，获得了丰富的考察资料，并做过了大量的实验，经过他反复思考和分析，形成了卓越的进化理论，这些理论集中发表在他的《物种起源》一书中。达尔文进化学说继承和发展了拉马克进化思想，他接受了拉马克的物种可变论，他认为进化是逐渐的和连续的。此外，达尔文还提出了新的观点，即共同祖先和自然选择的学说。达尔文认为进化性变化，并非拉马克式的神奇动力所导致的结果，也不是简单的偶然事件，而是选择的结果。自然选择学说被认为是达尔文进化论的核心。继《物种起源》出版后，1866年又发表了《动物和植物在家养下的变异》一书，1876年又出版了《植物异花受粉和自花受粉的效果》等著作，进一步充实了生物进化的理论。

随着科学的发展，特别是遗传学的建立，对达尔文进化学说进行了充实和科学论证，现在我们将要介绍的达尔文进化学说已是由许多学者根据染色体学说、群体遗传学、物种生物学概论和古生物学等所提出的现代新达尔文主义的观点。

(1)物种在不断演变，他们都起始于共同的祖先　地球上现存的物种，虽千差万别，但都是过去生存的物种的后代，都渊源于同一祖先。这种概念是很重要的，只有起源于共同的祖先，才能使生物界成为历史连续的统一整体，进化论才能建立在科学的基础上。百年来共同祖先作为阐明进化事实的基本概念，经过辩论终于得到了公认。现代遗传学的发展，证明一切生物(从病毒到人类)，都使用共同的遗传密码，这是在分子水平上提出的新的证据。

(2)自然选择　自然选择是达尔文理论的重心。在这个理论中"适者生存、不适者淘汰"是进化现象的驱动力。究竟自然选择的单位是群体、物种、个体、细胞、还是基因?1976年英国的R. Dawkins认为"自然选择以基因而非个体为单位",并称基因是自私的,因为它对于生物个体作贡献纯然是在长远上对自己的有利的行为。基因本身在生物个体之外是没有生存价值的,一个好的基因能使它们在以个体繁衍,它自己的繁衍依赖于个体的成败。

自私基因:Dawkins虽然是第一个把"自私基因"这个用词推广到一般用语里,但它们涵括的基因不见得是自私的。W. F. Doolittle及C. Sapienze给自私基因注入一个新的注解。自私基因是"无表现型,唯一功能是在个体的基因组内自我保存及繁殖的基因",换句话说,自私基因没有帮助个体的生存但是能够比基因组里其他基因复制得更快。例如转座元件(transposable element),简称TE。TE是散在于基因组里的小段DNA,这些DNA是可以在基因组里"转换位置"的基因。TE可以说是"寄生基因",寄生于基因组里面,偶尔它们会转座。它们通过两种办法繁衍增殖:跟宿主的复制而繁殖;在宿主的基因组内繁殖,即使宿主本身并不复制。这第二个办法是自私基因有别于其他基因的地方,它们的数量可以在宿主的基因组里迅速增长。

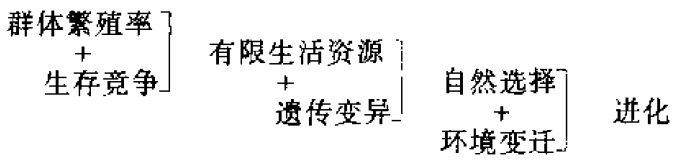
超自私基因:上述的自私基因虽然对于个体的适应值没有一点帮助,但大致也没有什么大害。一个成功的寄生虫是不会毁灭它的宿主的,也不会使宿主拖累得无法为它提供一个寄生处所。但还有另一类特别自私的自私基因,它们是靠破坏基因组里其他基因而从中获益的。Hammer于1990年称这类型基因为"超自私基因"。

这里仅举一个超自私基因的例子,即在*Nasonia*属黄蜂里的"父系性比"(paternal sexratio)染色体,简称Psr。Werren称Psr为"最狠毒的自私基因",它把同一基因组里所有的基因全部消除掉。我们先了解黄蜂及很多膜翅类昆虫的繁殖方法。在这些昆虫里,如果卵子经过受精,那么子代产生出来是雌性的,而没有受精的卵子产生的是雄性子代。所以雌性昆虫在这一目里是双倍体,就如同人一样;而雄性体是单倍体,就如同卵巢里的卵子一般。Psr是一个很特别的B染色体。Psr在雄性生殖细胞里发挥它的特别功能,于是我们可以假定它仅能在雄性体内发挥作用。Psr在下一代里非得进入雌性体不可,因为Psr只由精子携带而受精的卵子只发育成雌性子代。Psr把同一精子里其他染色体全部破坏,所以卵子虽然受了精,受精卵居然是单倍体。Psr就可以在这个单倍体受精卵里继续发挥它在雄性体里的特异自私效果。所以Psr在每一代都必须把跟它共处一代的所有其他染色体全部消灭掉;这些染色体(其实是染色体上的基因)提供Psr一个黄蜂个体所需要的所有功能,Psr只是坐享其成,到头来把这些基因扔到"进化的垃圾箱"里。

(3)逐渐进化　达尔文说过:"自然选择只能通过积累轻微的、连续的、有益的变异而发生作用,所以不能产生巨大的或突然的变化。它只能通过长而慢的步骤发生作用"。但这种观点曾遭到不少人的反对,连自封为"达尔文的斗犬"的赫胥黎也不能接受高等生物和新种的渐进起源说。随着科学的发展和对生物界更广泛深入的认识,渐变论得到了越来越多人的支持,并认为它是进化过程中最常见的最基本的形式。

骤变论之争也一直在持续着。它也有不少的证据,像异源多倍植物就是通过远缘杂交经染色体加倍而突然出现的新物种。

最后把达尔文的“通过自然选择的进化学说”的要点图解如下：



3. 非达尔文进化理论——中性学说

1968 年日本群体遗传学家木村(Kimura)和太田(Ohta)几乎同时提出分子进化中性论学说。中性学说又叫中性突变随机漂变学说。该系统较圆满地解释了生物大分子进化的两个特点：生物大分子层次上的进化改变不是由自然选择作用于有利突变而引起，而是在连续的突变压之下由选择中性或非常接近中性的突变的随机固定造成的。美国科学家 J. King 和 T. Jukes 发表了“达尔文主义进化”一文，他们支持木村的中性学说。这个学说是根据核酸、蛋白质中的核苷酸、氨基酸的置换速率，以及这样的置换造成核酸、蛋白质改变并不影响生物大分子功能的事实，提出了“中性突变”的概念。他们认为进化是“中性突变”在自然群体随机的遗传漂变的结果。这个学说对以自然选择为基础的达尔文主义进化论提出了新的挑战。

这个学说的要点如下：

(1)突变大半是“中性”　这种突变不影响核酸、蛋白质的功能，对个体生存既没有什么害处也没有什么好处，选择对它们没有作用。中性突变如同同义突变、同功能突变(蛋白质存在多种类型，如同工酶)、非同功性突变(没有功能的 DNA 序列发生突变，如高度重复序列中的核苷酸置换和基因间的 DNA 序列的置换)。这种中性突变由于没有选择的压力，它们在基因库里漂动，通过随机遗传移到群体中固定下来。

(2)分子进化的主角是中性突变而不是有利突变　中性突变率，也就是核苷酸和氨基酸的代换率是恒定的。细胞色素 c 中氨基酸的代换率在各种生物中也差不多相同的。所以蛋白质的进化表现与时间呈直线关系，可根据不同物种同一蛋白质分子的差别来估计物种进化的历史，推测生物的系统发育。这和化石以及其他来源推导出的进化关系是相符合的。此外还可根据恒定的蛋白质中氨基酸的替换速度，对不同系统发育事件的实际年代作出大致的估计，即所谓进化分子钟(molecular clock)。

(3)中性突变的进化是通过遗传漂移来进行的　遗传漂移使中性突变在群体中依靠机会自由结合，并在群体中传播，从而推动物种进化。所以生物进化是偶然的、随机的。

(4)中性突变分子进化是由分子本身的突变率来决定的　不是由选择压力造成的，所以分子进化与环境无关。

中性学说是在研究分子进化的基础上提出来的，用随机出现的中性突变，能很好地说明核酸、蛋白质等大分子的非适应性的多态性，认为根据核酸、蛋白质分子一级结构上的变化就可以说明生物性状的所有变异，进而说明进化原因。它否定了自然选择在进过程中的作用。

中性学说在学术界引起了激烈的争论，主要的反对意见是这种进化的偶然性成分太

大，任何一个物种或个体的出现都成了不可预料的偶然事件；对是否有真正的“中性”突变也也持有不同的看法。总之，这一理论吸引了很多人的注意，也有许多检验这一理论的实验。尽管目前还有不少争论问题，还不能证明它的精确性，但它已成为分子进化论中的一个有活力的部分。

二、进化的机制

根据达尔文的进化理论，生物进化的基本过程是自然选择，选择是在变异的基础上进行的。自然界的群体总是受经常发生突变和自然选择这两个重要因素的影响，正是它们不断地发挥作用，才促使群体的遗传组织发生变化。这种动态的演变显示着进化的过程。

进化的机制包含着遗传变异的产生和自然选择两方面的综合作用，在这种综合作用的过程中，我们从前面的章节中已了解了突变是怎样产生的，也知道了群体中遗传组成是怎样变化的，并能从量上给以估计。这两方面已经概括了进化发生和发展的机制。为了对进化过程作全面的理解，在这里补充说明两个问题。

1. 大量遗传变异的来源

遗传变异的产生除了基因的结构发生改变外，还包括基因的数量和组织结构的改变。由数百个个体组成的群体，每代可能只有几个突变，能引起生物特性发生重大的突变，往往不太可能是有利的。大多数进化性的改变似乎都是由于较小的突变的逐渐积累而出现的，这些较小突变伴有群体中个体特性的缓慢变化。突变尽管是罕见的，但它是遗传变异的源泉，是在点点滴滴为基因库增添新的等位基因，这是变异来源的一个重要方面。

遗传变异来源的另一个不可忽视的方面就是重组。基因由于重组而产生新的组合，在群体中虽然基因频率不改变，但重组和随机分配改组了群体中存在的基因，因而使等位基因的新组合每一代中都受到选择。即使在不增加突变的情况下，仅重组就可使群体在许多世代内暴露其隐藏的变异。重组能产生大量的遗传多样性，为群体提供了对变化着的环境作用适应的能力，大大增加了进化的可能性。

个体的杂合性是重组的基础，因为杂合性蕴藏着丰富的可重组性。个体的显性等位基因在大多数基因座上不是纯合的，群体中个体杂合性越大，在生殖细胞后代中等位基因可能组合的数量也越大。例如具有平均杂合性为 6.7% 的人体，假设有 100 000 个基因，则一个人有 67 000 种基因型是杂合的，这个个体就能产生 $2^{7000}(10^{17})$ 种不同的生殖细胞。其实这么多的生殖细胞绝非一个个体所能产生的，可以说全人类一下子也产生不了。由此可见，不会有两个人在遗传上是完全相同的，这就是人类个性的遗传基础。总之，群体中惊人的遗传变异量给进化的发生提供了大量的机会。

实验室的实验已证明，群体中遗传变异的数量越大，其进化的速度也就越快。在一次实验中，饲养两个果蝇群体，使一个群体最初具有的遗传变异约为另一群体的二倍。然后在实验中，让这两个群体在能促进迅速进化变异的条件下（处于对食物和生活空间紧张竞争的状态下）繁殖 25 代。虽然两个群体都繁殖起来了，并逐渐变得能较好适应实验室环境，但是在具有较多初始变异的群体中，进化的速率却高得多。

2. 如何保持与变化的环境相对应的巨大遗传变异库

群体中储存着大量的等位基因，在自然选择的作用下，有利的等位基因保存下来，不利的等位基因被淘汰。例如有一系列复等位基因，由于自然选择的结果，一个等位基因的频率是靠牺牲另一个或几个等位基因而增加的，直到这个基因座成为纯合的。这样，群体就会在自然选择的作用下趋向于纯合，使丰富的遗传变异库趋于单调。

选择对稳性基因不利时，它们总是在杂合体的掩护下低频率地被保存下来，直到环境发生变化使它成为适应者为止。在一般情况下，杂合体具有优势，这样往往使不利基因以较高的频率保存下来。最为明显的例子便是镰状细胞贫血，这种病盛行于非洲赤道和中东一带。镰状细胞贫血患者通常在生育年龄以前就死亡，应该是引起这种疾病的稳性基因 s 越来越少，s 基因的频率越来越低。但是在非洲某些地区，s 基因的频率高达 30%～40%。这是因为具有镰状细胞基因的杂合体对致死的疟疾具有抵抗力，而正常的纯合体则没有抵抗力。显然杂合个体优于两种纯合体，它既免患镰状细胞贫血病，又不怕疟疾的侵袭。结果在群体中杂合的个体便优先成活和繁殖，镰状细胞等位基因便在群体中以较高的频率保持着。可见杂合体能够积极保持多样性而不顾选择的作用，把对当时当地不利的基因淘汰掉。群体的特点就是杂合性，杂合具有优越性，它能适应较为广泛的环境条件，还可使杂合体开辟边缘生态环境。

在群体中，选择也可能直接地起保存复等位基因的作用。如果一个物种的生活区域包含有几种不同环境的话，那么自然选择就可使基因库发生多样化。其途径是使若干个等位基因能最好地适应于不同的小环境。通过这样的多样化选择，使某些基因座上保持复等位基因，从而保持了群体中的多态现象。

群体中保持复等位基因还有一种可能的机制，就是频率相依选择，即两个等位基因的适应性不是恒定的，而是随着它们的频率而改变的。如果一个等位基因在频率高时较另一个等位基因不利，那么当其频率下降到某一水平时，它就可获得有利性，于是这个等位基因的频率就会在那个水平上稳定下来。

三、物种的形成

在有性繁殖的生物体中，物种内是能相互交配的，而与另一物种的成员则不能相交配，这就是区别种的标准。一个种群内的个体之间，在遗传上也是千差万别的，但这种种内的差异在遗传上是连续的，它们之间可以进行基因交换，而种间的差异就不同了，种间的变异在表型上和遗传上都是不连续的，它们之间不能交配，不能交换基因，它们之间有着某种隔离机制(isolation mechanism)。如果没有隔离，群体内任何个体所发生的突变，如果选择上有利，就会逐渐扩散到整个群体。不同个体所发生的突变在群体中可以进行各种组合，一个群体始终保持一个群体。如果有隔离，就把群体分成为两个，自然选择必然对两个群体独立地起作用，一个群体中所发生的突变不会扩散到另一个群体，这两个群体进行着彼此独立的进化。这样就造成了表型上的不连续，也就成为分类学工作者们所认为的"种"。总之，隔离使群体分化，达到新种的形成。

前合子生殖隔离有 5 个主要类型：

(1)生态隔离　即群体虽占居同一地区，但生活在不同地区，因此不能相遇。

(2)时间隔离　即动物的配偶和植物的开花发生在不同的季节或一天中的不同时间。

(3)个性生态隔离　即雄性和雌性之间性吸引微弱或缺乏。

(4)机械隔离　即动物的交配和植物花粉的传播，由于生殖器的形状或大小不同或花器的结构不同而受到阻碍。

(5)配子不能互相吸引　雄性动物的精子在雌性生殖道中也可以不成活或者花粉在柱头上不萌发。

后合子隔离机制有3种主要类型。

(1)杂种不适　即杂种合子不能发展或至少不能达到性成熟。

(2)杂种不育　即杂种不能产生有性功能配子。

(3)杂种衰退　即杂种后代的生存能力和可育性减退。

由于物种是在生殖上被隔离的群体，所以物种是怎样形成的问题实际上就是生殖隔离是怎样形成的问题。

物种形成一般有两个阶段，第一阶段要求两个群体间的基因交换中断，其方式通常是地理上的隔离。地理上隔离使它们不能进行基因交换，由于各自对当地环境条件或生活方式的适应，使它们在遗传上发生趋异。当被隔离的群体在遗传上变得更加不同时，它们之间便可能出现后合子隔离机制，因为杂种会具有不协调的遗传组成，导致生存力和可育性的减退。

物种形成的第一阶段通常是一渐进的过程，因此要决定两个群体是否进入物种形成阶段，往往是困难的。况且，第一联合体又是可逆的，如果在地理上隔离了一段时间后，两个群体又开始相聚，在杂种适应性损失不太大时，这两个群体就可能融合为一个群体。另一方面，如果杂交繁殖的后代，其生存力和可育性大大减退的活群体便有可能经历到物种形成的第二阶段。

第二阶段是前合子隔离机制的发展。这是由自杀选择直接促进的过程。后合子隔离机制显然浪费了大量的生殖力。如果产生一个不成活的杂种合子，就浪费了两个配子；如果杂种能活而不育，那么浪费的不仅仅是配子，还包括杂种在其生长期间所消耗的资源。虽然在前合子隔离中的配子隔离也是浪费配子，以及在不成功的求偶或失败的交配中也浪费能量，但总的来说，前合子隔离机制的浪费比后合子隔离机制的浪费少得多。因此，当已被后合子生殖隔离机制的两个群体相接触，自然选择便会迅速地促进前合子隔离机制的发展。让我们考虑一个简单的情况，假定某两个等位基因 A 和 a，A 有利于群体内交配，a 有利于与其它群体交配。如果后合子隔离机制在两个群体之间起作用，在具有正常适应性的后代中A是常见的，而在低适应性的杂种后代中 a 是普遍的。结果等位基因 a 的频率一代代减少。自然选择有利于前合子隔离机制的发展，以避免杂种合子的形成。这样两个群体在生殖上就被完全隔离开来，成了两个不同的物种。

第七节　展　　望[47~48]

人类的进化，至少在近几千年内发展很快，而文化的改变比生物学的改变更为显著。

文化的适应方式比生物学的适应方式要更为有效和多变。但是超有机进化并不能取消有机的进化，生物的进化在人类中继续着，而且它也许会采取较快的步伐，因为它受到了文化进化所从未有的那种刺激。文化的和生物的进化是相关联的。人类文化有可能存在和发展，就是由于人类文化的遗传基础持续着或改进着。同时，文化进化无疑是环境的最重要的源泉，它也促进了人类的生物进化。

一、人类进化

原中国科学报(1995·06)等几家报纸曾先后刊载了美国科学家关于人类未来命运的"科学论断"——人类进化趋于停止的消息。生物学知识告诉我们，一个物种若停止了进化，就意味着它行将灭绝。我们知到人类也是动物的一种，如果停止了进化，也将离绝种的时间不远了。

人类进化停止论者是通晓现代生物理论的，他们选择了在生物进化过程中，两个关键问题来论证他们的命题。首先，他们援引发生在被称为人类起源摇篮的非洲裂谷区的古人类进化的科学事实。500万年前，全球气候发生了激烈变化，气温降低引起森林面积大幅度减少，取而代之的是广大的稀树草原，原来生活在森林中的古猿人，被迫在草原上生活。为了生存，它们不得不去适应草原上生活环境，经过漫长的历史磨难，它们学会了用双脚在地面上跨步面行，学会了采集植物的花果根叶和捕猎动物；创造了工具和语言，最终变成了人。接着人类进化停止论者提出，现代人是生活在人类为自身创造的"人造环境"之中，因为自然环境对人类的压力已不起作用了，自然选择作用被废止了，人类进化失去了原动力面不再进化了。

生物在漫长的进化过程中，一直经受地球上自然条件不停变化的压力。自然选择无时不在地对生物进化起作用。例如在地球历史的太古代(20多亿年前)地球的大气层中是缺乏氧气的，那时的空气是还原性的，地球上的生物是厌氧气的。后来地球上进化出光合生物，它们产生的氧气，使大气层中含氧量增多，并且在高空形成臭氧层。由于臭氧能吸收太阳光中的紫外光线，消除了阳光的杀伤力，保护了地球上的生物，从而大大改善了生物的生活条件。后来的生物在这种由于生物活动而形成的新环境中，不光没有停止进化，反而大大加速了进化，生物更加繁荣。又如，地球上的陆地表面本来是无土壤的，后来由于低级生物的活动，地表面才有了土壤的，为后来的植物和生活于土里的动物和微生物创造了新的生活环境。在这新的生活条件下，陆生植物才大大地发展起来，并为陆生动物提供了生活条件，多种多样的陆生物动物因此兴盛起来。由于生物活动而改良了的环境，是不会使生物停止进化的。同理，由人类劳动而改善了的环境，也不会使人类停止进化。在人类发展史上，对人类生活影响最大的是农业，有了农业，人的生活才有了保障，衣食才有了长久稳定来源；有了农业，人才定居下来，盖房造屋，免受风雨、寒暑以及野兽的侵扰，人类才生活在舒适的"人造环境中"。历史证明因农业改善了生活条件并没有使人类趋向灭绝，相反，农业使世界人口大增。

人类进化停止论者提出，现代人类社会的人类群体间交流广泛，有互相融合的趋势，已不像古代那样，人类群体间由于地理或风俗习惯造成的互相隔离状态。人类群体之间隔离的消失，阻止了新的物种(超人)的形成，人类的历史将转入进化停滞时期。从理论上

看，这一论点是符合进化论中的边域成种理论的。按照这一物种形成方式，新的物种形成必须有隔离机制才行，即正在形成新种的居群和原种的祖先居群间的基因交流必须断绝，新种才能形成。但是我们不能忘记，物种形成还有另一种方式，即世系成种方式，这一方式无需隔离机制，新种和原种也无必要在地理上分开，即可形成新种。因此，美国人类进化停止论者的推断有失偏颇，是不妥当的。

从现在世界人口实际情况来看，尽管现代交通发达，人们的经济生活水准高了，全世界各民族人员交往多了，基因交流机会多了，但现今大多数民族中外来的基因成分并不很多，离全球人类混合到不分彼此的程度还相去很远；再说，地球是巨大的，世界各地自然条件千差万别，只要时间足够长，由于地域的环境条件不同，生物必然会发生分化，地理分布范围广的物种常具有多形现象便是例证。根据古生物学资料，在历史上，一个新的物种的形成，一般要经历一百年之久，新种的生存时间可达几十或上百万年之久。已知最古老的人科化石，非洲的南方古猿（与现代人种同科不同属）发现于 200～340 多万年前的地层中；直立人（如北京中国猿人，与现代人种同属不同种）化石最早记录约有 170 多万年；智人（即现代人种）的历史不过 20～30 万年的历史，而现在地球上生存的人类，是智人人种的一亚种，其历史仅有几万年。从古生物学家眼光看来，现代人种的历史还很短，与古生物种的寿命相比只不过是新出现不久的物种。

从生物学规律来看，人类既然是一个生物物种，就有其起源、发展和灭亡的过程，这是自然规律所决定的，是不依人的意志来决定的。人类进化历史和一般物种的进化史一样，也应是漫长的，远不是一个人的个体生命可比的，绝不是经过千百代人之后就能见端倪的。生物学上，物种绝灭有两种形式，一种形式是整个物种的全部生物个体都死亡，不留后裔；另一种形式，是随着时间的推移，原来的物种虽然全部生物个体都灭亡了，但它们遗留下了新型的后代生物（新种）。如果生物进化的历史进程很长，它的发展必然包含有许多随机因素，因此预言未来的历史命运是极难做到的。

二、生物进化工程

1973 年美国斯坦福大学的 H. Boyer 和 S. Cohen 采用 DNA 重组技术，首次获得外源 DNA 的分子克隆，从而开始了在分子水平上对遗传物质的加工和改造，即基因工程。10 年之后，K. M. Ulmer 提出了蛋白质工程的设想，即通过对蛋白质已知晶体结构的了解，借助计算机辅助设计，用基因定位诱变等技术改造基因，以达到改进蛋白质某些性质的目的。基因及其产物蛋白质都是生物在长期进化过程中形成的。基因工程和蛋白质工程为研究这些生物大分子结构与功能的进化提供了强有力的手段，在此基础上发展出了生物进化工程。所谓生物进化工程，就是通过模似生物大分子进化的过程，来创造新的基因、新的蛋白质和新的物种，这是一种定向进化的方法。在人工控制的条件下，使产物沿一定的方向演变，以产生符合人们的需要的特性分子，从而得以在短时间内实现自然界几百万年才能完成的生物进化过程。

1. 早期的生物分子进化实验

达尔文进化基本上包括三个过程的重复运行，即选择、增殖和突变。选择，无论是自

然界选择或人工选择，都是把“优者”从“劣者”中分离出来的筛选过程。自然界“优者”是指生活和生殖能力强的物质个体。实验室里“优胜劣汰”是按人们的需要来取舍的，可能并不符合生物自身发展的要求。分子选择也一样，只要有适当的方法就能将所需要的分子挑选出来。

在20世纪60年代后期，S. Spiegelman及其合作者就在实验室内证明了分子的达尔文进化。Spiogelman曾研究一种称为Qβ的RNA噬菌体。如将此病毒RNA与其复制酶混合并加入40种合成RNA的底物核苷三磷酸(NTP)，病毒RNA就大量复制。RNA复制酶很容易产生差错，因此在RNA复制过程中将会积累变异。他让Qβ在试管中复制20分钟，然后取出一定量反应混合物转移到另一试管中，补充酶和底物后继续复制，如此重复74次。他还利用缩短复制时间给选择施加压力，因此需要较长的复制时间的分子必然不能像需时短的分子那样留下众多的后代，产生的RNA分子会变得越来越短，因为较小的分子可以较快地复制。到第74次复制时，QβRNA已被删去了80%，只剩下RNA两端为复制酶识别并控制起始复制所必要的序列，它们将RNA复制的效率提高大约15倍。此过程揭示了RNA分子在试管内的进化过程，然而其应用却受到严重的限制，因为Qβ复制酶对于被复制的RNA起始端序列十分苛求。

2. PCR技术用于DNA的定向进化

聚合酶链式反应(PCR)是目前生物技术中应用最广的DNA体外扩增法。该方法的发明者K. B. Mullis因此而得到1993年诺贝尔化学奖的殊荣。PCR技术是DNA通过重复变性、退火和延伸步骤，使DNA基因扩增数百万倍。科学家们已将PCR技术用于DNA的定向进化，从中选择出一些具有特殊性能的DNA分子。

目前已有一些能溶解血栓的蛋白质药物，可用来治疗心机梗塞、脑血栓等疾病。遗憾的是，病人可能对外源蛋白质产生过敏反应，于是科学家们设想利用不易产生过敏反应的DNA来开发一种抗凝血药物。J. J. Toole及其同事不是向自然界去寻找药物，而决定通过定向进化来获得抑制凝血酶活性的DNA分子。他们构建了1013个核苷酸的单链DNA分子，各具有不同的序列，把这些DNA分子群体同固定化于固体支持物表面的凝血酶相接触。几乎绝大部分DNA分子都未与凝血酶结合而从支持物上被洗脱，但有极少数DNA分子仍然与凝血酶相结合。将这部分DNA分子回收，并用PCR技术扩增，然后再经固相化凝血酶选择。如此重复5个循环，最终得到高度富集的具结合凝血酶能力的DNA，从中分离到可以抑制血栓形成的药物分子。目前医生正将这些药物在狒狒和猴子身上作实验，这种药物表现出较强的抗凝血作用。

3. RNA的定向进化

定向进化必须将选择与增殖直接联系起来。目前比较通用的RNA体外扩增技术涉及两种聚合酶：一种是反转录酶，即以RNA为模板合成DNA的酶；一种是RNA聚合酶，它以DNA为模板合成RNA。在此转录-反转录系统中需要一对引物，其中一个引物附加有转录的控制系统(称为启动子)，另外还要加入合成RNA的4种dNTP和合成DNA的4种dNTP。该系统具有比PCR更高的扩增效率并且无须进行热循环。

定向进化系统一开始就要有足够的异质性，并在增殖过程中引入变异，否则选择就没

有对象。要达到建立异质群体的目的，有三种方式供选择：

(1)鸟枪法　使大分子的所有位点都可发生随机变异，在寻找新的特性分子时，此方法是合理的，但要选择的数目往往显得过大。

(2)覆盖法　使随机变异仅发生在所选定的区域内，此法需对大分子结构与功能有较多的了解，并对选择的目的十分明确和有限。

(3)集中法　当所希望获得的大分子是已知大分子的某种改进性能的突变体时，可用这种分子的一组突变体。此时与已知分子序列越接近的分子数目增大，差异大的分子数目逐渐减少。由于新的突变是在原有经选择的变异基础上发生的，因而具有倾向性。重复随机突变是在实验室内实现达尔文进化的关键。

在建立了增殖和突变的基础上就可以用各种选择方法去研究 RNA 的进化。G. F. Joyce 小组在这方面做了卓有成效的研究，并将 RNA 进化推向优化它们功能特性的方面。他们利用天然存在并可切割 RNA 的核酶 RNA(ribozyme)，经过定向进化，产生可切割 DNA 靶序列的 RNA 分子。这样一种类酶 RNA 可以作药物使用，例如，它可用来对付入侵的病毒 DNA，在目前还几乎没有防治病毒的特效药物。

他们在四膜虫的一种类酶 RNA 基础上建立了 10 个变异株作为起始群体。通过与 DNA 靶序列接触，获得能与其相结合的 RNA 分子，经过选择性增殖，并在增殖过程中引入新的突变以产生具有更强结合和分解能力的分子。将这一过程重复 10 次，就能产生有效分解 DNA 的新的 RNA 分子群体。

达尔文的自由选择、增殖和突变的法则已成为科学家手中有力的工具。一些实验室正利用它重建分子进化过程。在进化过程中，RNA 曾具有过一些酶的性质，如 ATP 酶活性，后来丢失了，现在科学家又重新使其获得这种性质，希望以此产生更多新的性质。

4. 噬菌体展示技术

蛋白质是细胞主要的功能分子，它的定向进化意义更大、更有实用价值。然而这里遇到了一个复杂的问题，我们希望得到的是蛋白质，增殖与突变却不是人选分子本身，而是它们的遗传基因。如何能将其统一起来？20 世纪 80 年代末发展起来的噬菌体展示(phage display)技术为此提供了可能的途径。

fd 和 M13 是一类含单链环状 DNA 的丝状噬菌体。其基因 *III* 编码吸附蛋白(pIII)，该蛋白 C 端附着于噬菌体末端，N 端伸展在外。将外源基因重组插入到基因 *III* 编码蛋白的 N 端附近，表达产物成为融合蛋白，外源蛋白以独立的空间结构伸展在 pIII 蛋白的 N 端。如果将外源基因经诱变构建一个突变库，然后把所有突变体基因插入噬菌体的基因 *III* 位点，突变体蛋白就能在噬菌体末端被展示。利用抗原-抗体反应或其他亲和反应吸附噬菌体可以选择出结合力强的外源蛋白及其基因。

噬菌体展示技术一出现立即受到广泛的重视，许多实验室用此技术去选择特定构象的蛋白质。例如，可以用它来得到特定抗体。对此，先要从人的淋巴细胞中取得 DNA。再用 PCR 技术扩增出全部抗体基因，然后将这些抗体基因插入噬菌体基因 *III* 的 N 端，收集成熟的噬菌体，通过固定化的抗原柱吸附带有特异抗体的噬菌体，经回收并增殖噬菌体后再次用亲和法纯化带有抗体的噬菌体，这样通过两轮选择就能分离到特异的抗体及其基因。噬菌体展示技术比单克隆抗体技术有更多的优点。由于制备单克隆抗体时只能

用动物来免疫抗原，所得到的抗体不是人体的抗体，这就限制它在临床上的应用；而噬菌体展示技术通过扩增人的抗体基因来选择出人的特异抗体，因此可作为药物直接用于临床治疗。噬菌体展示技术具有级高的选择效率，且易于操作。噬菌体展示技术还便于进行抗体工程，能产生单链抗体、双单链抗体、小抗体等改进性能的抗体。

噬菌体展示技术已用于从肽库和突变库中选择多种蛋白质。严格地说，这样一种过程不是进化性的，因为它没有为每一代中产生新的突变提供机遇。然而，一些更为复杂的实验系统则为每代都提供变异的机会。所以，选择、增殖和突变能够不断地沿着一个方向进行下去。

5. 生物进化工程有广阔的发展前景

生物进化工程采用定向进化技术来创造新的基因、新的蛋白质和新的物种。进化工程与定向进化具有相同的含义，但前者表明这是一类生物工程，后者是指这类工程的技术系统。生物进化工程目前还处于起始阶段，许多技术有待于改进和提高，还需要作更多的探索。但是它的问世已经受到普遍的关注。最明显的是医药界，已把今后开发新的治疗一些恶性疾病的药物，例如心血管系统疾病、癌症、艾滋病、老年性痴呆症和免疫障碍等疾病的特效药物，寄希望于这一途径。生物进化工程还可能被用来解决医学上的一些难题，因为疾病的产生也是与人类自身的进化过程有关的。利用生物进化工程来改良物种，也是另一个诱人的方面。生物进化工程还有着重大的理论意义。生物分子进化是一个十分重要的基础理论研究，它的成果可用于指导生物进化工程；另一方面生物进化工程又能在实验室揭示生物分子详细的进化历程。这里再次表明了基础理论研究与发展高新技术之间的密切关系。总之，生物进化工程虽然出现不久，却已表明它具有广泛的发展前景，这是一个充满希望、十分诱人的研究领域。

三、进化研究的新理论对生物进化工程的影响

1. 重组 DNA 技术与新类型生物形成

现代物种形成与进化理论指出：基因突变、染色体畸变、基因重组可改变物种的基因型，从而形成新性状。在这一理论指导下，先找出某种特定生物（植物、微生物）的基因，然后把这些基因移植到某一植物细胞中去，虽然这种技术难度较大，但从新的物种形成理论上讲，它是可能成为培育植物新品种的一种方法的。生产实践证明了这一点，现已培育出能生产杀虫的烟草和棉花的新品种。

2. 突变和物种新类型的创造

突变在细胞遗传学中主要指染色体的不连续变化，包括缺失、移位和倒位等。突变在物种形成作用的研究中，已直接为育种新技术奠定了理论基础。例如辐射育种和化学诱变育种，以及突变的筛选，都是生物工程中前沿的研究领域。如在玉米的诱发突变实验中，发现 X 射线引起的突变是由于基因损伤造成的。现已培育出优良的玉米新品种。

3. 多倍体物种形成和物种新类型合成

多倍体物种在植物界中广泛存在，植物界中多倍体起源的种类约占 30%～35%，一些重要的农作物如小麦、马铃薯、香蕉、甘蔗、燕麦、棉花、烟草、咖啡等都是多倍体。用现代技术形成的同源多倍体新种已用秋水仙碱和其他化学试剂使染色体加倍获得成功。用实验的方法已合成了几乎与现存天然种相同的多倍体种，如唇形科鼬瓣花属中“人造 Teirahit”，在形态学、遗传学和细胞学上都很像天然种；菊科 *Madia* 属中由 *M. gracilis* 和 *M. citriodora* 合成了与 *M. citrigracilis* 非常相似的种；在十字花科芸苔属中，芥菜可由黑芥（*B. nigrs*）和油茶（*B. campestris*）合成，蔓青（*B. napus*）可由甘蓝（*B. oleracea*）和油菜（*B. campestris*）合成。当今，用生物工程技术重新创造的农作物已有棉花、小麦和烟草等。

4. 体细胞无性繁殖育成新品种

运用该生物技术方法育种，是把细胞放在一些特别的激素和营养物中进行培养，这些细胞能显示出不同于一般细胞新的遗传变异，然后将一些符合人们需要或最有商业价值的变种分离出来，进一步进行培养研究，美国维吉斯纳克思公司采用这种方法已育成味美可口的番茄新品种。

5. 无融合生殖和物种新类型的获得

在研究生物进化过程中发现，某些植物在系统发育中性别会发生退化，而且会退化到失去基本的进化功能的地步，即从有性生殖退化到无融合生殖时、在进化过程中形成了一种特殊的“封闭式”状态，它们不能进化出任何新东西，而只能通过其中一些有性成员的活动与改变了的环境一并前进，这类物种的繁殖即称兼性列融合生殖。研究发现，在以多倍体方式形成物种时产生的多倍体中，无融合现象对某一组内染色体的重复或缺失而产生不稳定的非整倍体、单体和多体物起着保护和固定作用。无融合生殖产生了纯合体。无融合生殖已成为育种学家当今重点研究的前沿育种方法之一，因为选种方法能保持和固定已获取性状的固定，促使这物新物种的形成。

（谭　芳　陈守义）

参 考 文 献

[1] Levin S A et al. Ecological Genetics. New York: Springer Verlag, 1978

[2] Jukes T H et al. Evolution of Life. Tokyo: Spring Verlag, 1991

[3] Osawa S T H et al. Microbiological Reviews, 1992. 56: 229～264

[4] Okada N et al. J Mol Evol, 1993. 37: 167～170

[5] Tamura K et al. Mol Biol Evol, 1993. 10: 512～526

[6] Vigilant L et al. Science, 1991. 253: 1503～1507

[7] Gillespie J H. The cause of molecular evolution. New York: Oxford University Press, 1991

[8] Jeffreys AJ et al. Electrophoresis, 1995. 16: 1577～1585

[9] Vergnaud G et al. Genomics, 1991. 11:135～144
[10] Britten RJ et al. Science, 1969. 165:349～357
[11] Bachmann K et al. Biol, 1972. 23:419～450
[12] Turner SH et al. Biochem Gene, 1973. 10:263～274
[13] Ohno S et al. Sex chromosomes and sex-linked genes. Berlin: Springer, 1967
[14] Ohno S. Evolution by gene duplication. New York: Springer, 1970
[15] Richmond RC et al. Evolution, 1972. 26:536～544
[16] Waring M et al. Science, 1966. 154:791～794
[17] Hess O et al. Advance Genet, 1968. 14:171～223
[18] Ohta T. Genet Res, 1971. 18:277～286
[19] Crick F H C. Nature, 1971. 234:25～27
[20] Jeffreys A J et al. Nature, 1991. 354:204～209
[21] Shriver M D et al. Genomics, 1992. 14:449～454
[22] Go M et al. Nature, 1981. 291:90～92
[23] Gilbert W et al. Cell, 1986. 46:151～154
[24] Littman D R et al. Nature, 1987. 325:453～455
[25] Jaworski C J et al. Nature, 1989. 337:752～754
[26] Nagylaki T et al. Genetics, 1990. 126:261～276
[27] Symington L S et al. Mol Cell Biol, 1988. 8:595～604
[28] HammerM F et al. J Hum Genet, 1995. 56:951～962
[29] Baker C S et al. Proc Natl Acad Sci USA, 1993. 90:8239～8243
[30] Lewin R. Science, 1981. 214:426～429
[31] Hamby R et al. Molecular systematics in plants. New York: 1992
[32] Irwin D M et al. J Mol Evol, 1991. 32:128～144
[33] Peng C K et al. Nature, 1992. 356:168～170
[34] 李婧炎等。大自然探索，1993。12(4):16～21
[35] Zhang Y P et al. Proc Natl Acad Sci USA, 1993. 90:9557～9561
[36] Zhang Y P et al. Mol Phylogenet Evol, 1994 3:351～359
[37] 张尚宏等。动物学研究，1995。16(2):32～45
[38] Zhang S H. Speculat Sci Technol, 1996. 1:19～22
[39] 张尚宏等。中山大学学报，1996。35(5):96～101
[40] Palmer J D et ak. Molecular evolution genetics. New York: Plenum Press, 1985
[41] Cech T R et al. Cell, 1986. 44:207～210
[42] Nei M et al. Molecular evolutionary genetics. New York: Columbia Uviversity Press, 1987
[43] Wheeler W C et al. Mol Biol Evol, 1988. 5:90～96
[44] Gray M W. Annu Rev Cell Biol, 1989. 5:25～50
[45] Bork P et al. Biology, 1992. 2(3):413～420
[46] Bock L C et al. Nature, 1992. 355:564～568
[47] Beaudry A A et al. Science, 1992. 257:635～638
[48] Smith G P. Gene, 1993. 128:1～8

汉英对照索引

A

Alu 家族 *Alu* family 90,445
ARS 结合因子 1 *ARS*-binding factor 1, ABF1 148
ARS 共有序列 *ARS* consensus sequence, *ACS* 147
AU 元件 AU-rich element, ARE 182
a 型活性匣子 active cassette 367
癌基因 oncogene 432
氨甲蝶呤 methotrexate, MTX 378
氨酰 tRNA 合成酶 amino-tRNA synthetase 202

B

8-羟-5-脱氮核黄素 8-deazaribflavin 412
八聚核苷酸元件 octemer element 165
白细胞介素 7 interleukin IL 7 393
摆动假说 wobble hypothesis 200
斑驳现象 variegation 338
半保留复制 semiconservative replication 98
半不连续复制 semidiscontinuous replication 103
半翅目 Hemiptera 10
胞嘧啶 cytosine 52
胞液 cytosol 5
保守型转座 consertive transposition 328
比连接差 specific linking difference 68
臂间倒位 pericentric inversion 40
臂内倒位 paracentric inversion 40
编码区 coding region 74
变性 denaturation 58
表达位点 expression site 376
表达相连的拷贝 expression linked copy, *ELC* 376
表型 phenotype 72
病毒颗粒 virion 49
病毒类超家族 viral superfamily 353
病毒样颗粒 virus-like particle, VLP 356
补钉型重组 patch recombination 306
补骨脂素 psoralen 209
不等交换 unequal crossing over 40
不精确的切除 imprecise excision 330
不均一核 RNA heterogenous nuclear RNA, hnRNA 159
不可诱导 uninducible 234
不连续基因 interrupted gene 182
不稳定序列 slippery sequence 207

C

CAAT 框 CAAT box 165
cDNA 文库 cDNA library 92
CG 岛 CG island 93
Chargaff 法则 Chargaff's rule 48,52
C 带 C-banding 23
C 末端结构域 carboxy-terminal domain, CTD 159
C 值 C-value 22,82
C 值矛盾 C value paradox 82
C 分带法 C-banding 23
侧成分 lateral element 33
侧链假说 side-chain hypothesis 386
侧翼元件 flanking element 12
插入序列 insertion sequence, IS 324
缠绕数 writhing number 68
长散布元件 long interspersed element, *LINE* 90,353
长末端重复 long terminal repeat, LTR 349
常染色体 autosome 21
常染色质 euchromatin 20
操纵基因 operator 230,231
操纵子 operon 231
超二倍体 hyperploid 43
超基因家族 gene superfamily 88
超家族 superfamily 274
超螺旋 supercoil 67
超螺旋密度 superhelix density 68
超去阻抑 super-derepressed 244
糙面内质网 rough endoplasimic reticulum, rER 5
沉默基因 silencer 264,268
沉默匣子 silent cassettes 367
成核聚合 nucleation-polymerization 50
重复 duplication 442
重叠基因 overlapping gene 79
重合进化 coincidental evolution 454
重链, H 链 heavy chain 137
重组活性基因 *RAG-1*, *RAG-2* recombination activat

ing genes, *RAG-1*, *RAG-2* 393
重组节 recombination nodule, RN 34
重组期 recombination stage 34
重组小结 recombination nodes 298
重组修复 recombination-repair 422
赤道板 equatorial plane 30
串联重复 tandem repeat 62
次缢痕 secondary constriction 11
持家基因 housekeeping gene 265
雌激素 estrogen 273
从头起始 *de novo* initiation 117
粗糙脉胞菌 *Neurospora carassa* 35,194
粗线期 pachynema 34
促性腺激素释放激素 gonadotropin-releasing hormone, GnRH 383
催乳激素 prolactin 462

D

DNA 分析遗传学 genetics by DNA analysis 435
DNA 合成前期 pre-synthetic gap phase 27
DNA 聚合酶 DNA polymerase 102
DNA 聚合酶 I DNA polymerase I 107
DNA 聚合酶 II DNA polymerase II 107
DNA 聚合酶 III DNA polymerase III 106,108
DNA 聚合酶 δ polymerase δ 143,146
DNA 聚合酶 ε polymerase ε 143,146
DNA 连接酶 DNA ligase 113
DNA 序列家族 DNA sequence family 87
DNA 旋转酶 DNA gyrase 69,116
大沟 wide groove 55
大肠菌素 colicin 212
单倍体性 haploidy 42
单点突变 point mutation 428
单核苷酸多态性 single nucleotide polymorphism, SNP 96
单链断裂 single-strand break, SSB 406
单链交换 single-strand exchange 423
单链结合蛋白 single-strand binding protein, SSB 116
单链摄取 single-strand uptake 305
单链同化 single-strand assimilation 305
单体 monosomy 43
蛋白激酶 C protein kinase C, PK-C 287
倒位 inversion 39,441
灯刷染色体 lampbrush chromosome 18
等臂染色体 isochromosome 42
等位基因 allele 439
等位基因间的互补 interallelic complementation 236
第二附着点 secondary attachment site 311
颠换 transversion 428
定位因子 positoning factor 174
动基体 kinetoplast 457
动力学模型 dynamic model 281
动粒 kinetochore 9
动态突变 dynamic mutations 318
端化作用 terminalization 35
端粒 telomere 11,62
端粒结合蛋白 telomere binding protein 12
端粒酶 telomerase 151
端着丝粒 telocentric 10
短散布元件 short interspersed element, *SINE* 90,354
断裂-融合-桥 breakage-fusion-bridge, BFB 314
多倍体 polyploid 43
多点突变 multiple mutation 428
多复制子 multireplicon 142
多拷贝抑制 multicopy inhibition 336
多态性 polymorphism 96
多线染色体 polytene chromosome 19
多着丝粒 polycentromere 10

E

二步过程假说 two-step process hypothesis 35
二氢叶酸还原酶 dihydrofolate reductase, DHFR 378

F

发夹 hairpin 59
发夹环 hairpin loop 60
翻译 translation 199
翻译扩增 translational amplification 287
翻译区域 translation domain 210
反带法 reverse banding, R-banding 23
反接重复 reverse duplication 40
反密码子 anticodon 200
反式剪接 *trans*-splicing 284
反式显性 *trans*-dominant 236
反式作用 *trans*-acting 229
反式作用因子 *trans* acting factor 229
反求遗传学 reverse genetics 434
反向末端重复 inverted terminal repeats 324
反向重复 inverted repeat 61
反转座子 retroposons 345
范德瓦耳斯力 van der Waals force 55
范可尼贫血 Fanconi's anemia, FA 426

纺锤体 spindle 12
非保守性替代 nonconservative substitution 429
非病毒类超家族 nonviral superfamily 353
非重复 DNA nonrepetitive DNA 86
非复制型转座 nonreplicative transposition 328
非条件性突变 nonconditional mutation 429
非整倍性(变异) aneuploidy 43
非自主元件 nonautonomous element 339
分解代谢产物阻抑作用 catabolite repression 246
分解代谢物激活蛋白 catabolite activator protein, CAP 246
分支酸 chorismic acid 258
分支位点 branching site 185
分子团 micelle 4
分子钟 molecular clock 466
佛波酯 TPA 12-O-tetradecanoyl phorbol-13-acetate 287
辅阻抑物 corepressor 233,244
辅助病毒 helper virus 352
负超螺旋 negative supercoil 68
负互补作用 negative complementation 236
负链 DNA minus strand DNA 349
负异固缩现象 negative heteropycnosis 21
附着点 attachment site, att 311
父系性比 paternal sexratio 466
复合型转座 composite transposition 328
复合易位 complex translocation 41
复性 renaturation 58
复性动力学 reassociation kinetics 84
复杂性 complexity 84
复制 replication 98
复制叉 replication fork 101
复制蛋白 A replication protein A, RP-A、RF-A 145
复制倒位 duplicative inversions 329
复制酶 replicase 107
复制缺陷型 replication-defective 352
复制体 replisome 121
复制型转座 replicative transposition 328
复制眼 replicative eye 101
复制因子 C replication factor C, RFC 146
复制终止蛋白 replication terminator protein, RTP 130
复制子 replicon 100

G

GC 框 GC box 165
G_1 期 G_1 phase 27
G_2 期 G_2 phase 29
G-带 G-banding 22
γ复合物 γ-complex 108, 121
冈崎片段 Okazaki fragment 103
高度重复 DNA 序列 highly repetitive DNA 86
高尔基复合体 Golgi complex 5
高尔基器 Golgi apparatus 5
高分辨显带法 high-resolution banding 24
割裂基因 split gene 76,182
隔离机制 isolation mechanism 468
共合体 cointegrate 332
共价延伸 covalent extension 117
共有序列 consensus sequence 313
古细菌 *archaebacteria* 3
关连 tRNA cognate tRNA 202
归巢 DNA 内切酶 homing DNA endonucleases 297
滚环复制 rolling circle replication 141
果蝇 *Drosophila* 7

H

Holliday 连接体 Holliday junction 293,295,300
含氮碱基 nitrogenous base 52
合子 zygote 43
核分裂 karyokinesis 29
核苷 nucleoside 53
核苷酸 nucleotide 52
核苷酸切除修复 neucleotide excise repair, NER 410
核基质 nucleoplasm 8
核孔 nuclear pore 8
核酶 ribozyme 473
核膜 nuclear envelope 8
核仁 nucleolus 8
核仁组织区 nucleolus organizing region, NOR 8
核素 nuclein 47
核酸 nucleic acid 47
核糖体 ribosome 207
核糖体结合位点 ribosome binding site 213
核糖体释放因子 ribosome releasing factor, RRF 224
核小体 nucleosome 13
核小体核心 nucleosome core 14
核心 DNA core DNA 14
核心段 core 239
核心酶 core polymerase 155
核心启动子 core promoter 164
核心序列 core sequence 311
核心聚合酶 polymerase 107 核型 karyotype 25

核型图 karyogram 26
黑腹果蝇 *Drosophila melanogaster* 268,441
亨廷顿氏病 Hungtington's disease ,HD 318
恒定区 constant region,C 387
后复制修复 post-replication repair 421
后期 anaphase 31
后期延迟 anaphase lag 44
后随链 lagging strand 104
互补测验 complementation test 235
互补决定区 complementarity-determining region ,CDR 388
互补群 complementation group 235
花束期 bouquet stage 33
滑动钳 sliding clamp 107, 121
滑面内质网 smooth endoplasmic reticulum,sER 5
环结构域 loop domains 16
黄色霉素 kirromycin 216
磺酸乙基甲烷 ethyl methane sulfonate,EMS 407
回复过程 recover process 409
回复突变 reverse/back mutation 429
回文序列 palindrome 61
混倍体变异 mixoploidy 44
活性染色质 active chromatin 20

J

J 片段 Joining segment, *J* 390
交换固定 crossover fixation 447
校正 proofreading 109
肌动蛋白 actin 285
肌红蛋白 myoglobin 460
肌萎缩性(脊髓)侧索硬化症 amyotrophic lateral sclerosis 426
基本环化反应 primary cyclization 194
基本拷贝基因 basic-copy gene 377
基础水平元件 basal level element,BLE 268
基础转录因子 basal transcription factor 171
基础转录装置 basal transcription apparatus 171
基因 gene 72
基因簇 gene cluster 87
基因家族 gene family 87
基因内回复 intragenic reversion 429
基因敲除 knocked out 384,393
基因突变 gene mutation 427,439
基因型 genotype 72,384
基因转换 gene conversion 34,301,447
基因组 genome 81
基质 cytoplasmic matrix 5
激活物 activator 244
吉姆萨分带法 G-banding 22
剂量效应 dosage effect 40
假基因 pseudogene 87,358
假显性现象 pseudo dominance 40
甲基化不足 undermethylation 281
间带 interband 19
间隔区 spacer region 75
间期 interphase 27
兼性异染色质 facultative heterochromatin 21
减数分裂 meiosis 12,32,72,98
减数分裂后分离 postmeiotic segregation 301
剪接 splicing 77
剪接前导序列 spliced leader,SL 284
剪接位点 splicing sites 184
剪接型重组 splice recombination 308
剪接增强子 splicing enhancer 189
简并 degeneracy 200
简单易位 simple translocation 41
碱基插入 base insertion 428
碱基堆积力 base stacking 55
碱基缺失 base deletion 428
碱基替代 base substitution 429
鉴别碱基 discriminator base 204
交换 crossing over 33,72
结构基因 structure gene 74,229
姐妹染色单体互换 sister chromatid exchange,SCR 24
解离酶 resolvase 328
金属硫蛋白 metallothionein,MT 268
近端序列元件 proximal sequence element, PSE 167
近端着丝粒 subtelocentric,acrocentric 10
近中着丝粒 submetacentric 10
进化钟 evolutionary clock 453
聚合酶链式反应 polymerase chain reaction 439,472
均一化作用 homogenization 447
均匀染色区 homogeneously staining region,HSR 379

K

Klenow 片段 Klenow fragment 106
抗原变异 antigenic variation 375
抗终止作用 antitermination 177
拷贝选择 copy choice 349
克雅氏病 Creutzfeldt-Jakob disease,CJD 49
科克因综合征 Cockayne's syndrome 425
可变的表面糖蛋白 variable surface glycoprotin, VSG

374
可变的主要蛋白 variable major protein, VMP 378
可变剪接 alternative splicing 283
可变区 variable region, V 387
可读框 open reading frame, ORF 83
可移动的同源区 portable regions of homology 323
空转反应 idling reaction 257
控制元件 controlling elements 322
扣除杂交 subtractive hybridization 93
快停突变体 fast-stop mutant 105
喹吖因 quinacrine 22
扩增 amplification 378

L

类 V 结构域 V-like domain 452
类病毒 viroid 49
类核 nucleoid 4
类固醇受体 steroid receptor 269,272
利福霉素 rifamycin 155
连环分子 concatemer 132
连接 DNA linker DNA 14
连接数 linking number 68,308
连锁 linkage 72
连锁群 linkage group 72
联会 synapsis 32
联会复合体 synaptonemal complex, SC 33
镰状细胞贫血 sickle cell anemia 72
链滑动 strand-slippage 314
链溶霉素 streptolydigin 155
两亲性分子 amphipathic molecule 4
亮氨酸拉链 leucine zipper 270,278
磷酸二酯键 phosphodiester linkage 47,52
螺线体模型 solenoid model 16
螺旋-环-螺旋 helix-loop-helix, HLH 270,276
螺旋-转角-螺旋 helix-turn-helix, HTH 239,270

M

马蛔虫 *Ascaris megalocephala* 22
慢停突变体 slow-stop mutant 105
帽子结构 cap 180
帽子结合蛋白 cap binding protein, CBP 166
帽子位点 cap site 164
锚定蛋白 docking protein 226,402
酶的转换 enzyme turnover 310
孟德尔基因座 Mendelian locus 382
弥散着丝粒 holocentromeres 10
密码子 codon 74,200
密码子家族 codon family 200
嘧啶 pyrimidine 52
模板辅助 template-assisted 51
末端蛋白 terminal protein, TP 133
末端缺失 terminal deletion 39
末端冗余 terminal redundancy 132
末期 telophase 31

N

Northern 印迹 Northern blot 92
内部核糖体进入片段 internal ribosome entry segment, IRES 221
内部环 interior loop 60
内部启动子 internal promoter 166
内部引导序列 internal guide sequence, IGS 192
内含子 intron 3,76,184
内膜系统 internal membrane system 5
内网器 internal reticular apparatus 5
内源性原病毒 endogenous provirus 346
内在蛋白 intrinsic protein 5
内在终止子 intrinsic terminator 168
内质网 endoplasmic reticulum, ER 5
囊泡 coated vesicles 6
逆环化 reverse cyclization 194
念珠模型 beads-on-a-string 13
鸟嘌呤 guanine 52
鸟嘌呤四联体 G-quartet 61
尿嘧啶 uracil 52
凝胶阻滞法 band shift 168
凝线期 synizesis 33
扭转数 twisting number 68,308
纽体 ν-body 14

O

偶线期 zygonema 33

P

Polα/引发酶 polα/primase 146
poly(A)结合蛋白 poly(A)-binding protein, PABP 180
Pribnow 框 Pribnow box 161
PRP 基因 pre-RNA processing gene 191
胚胎干细胞 embryonic stem, ES 383
嘌呤碱 purine 52
剪接体 spliceosome 187
平行汇接 paranemic joint 306

Q

Q带 Q-banding 22
启动子 promoter 161,230
起始DNA initiator DNA,iDNA 146
起始延伸复合物 initial elongation complex 171
起始因子 initiation factor 213
起始子 initiator, Inr 166,265
前导链 leading strand 104
前导序列 leader sequence,L 284,402
前期 prophase 30
钳载体 clamp loader 108, 121
嵌合型 chimerism 44
强停止负链DNA strong stop minus strand DNA 349
强停止正链DNA strong stop plus strand DNA 349
羟基脲 hydroxyurea 29
切除 excise 311,329
切除酶 excisionase,Xis 312
切口 nick 141
切口移位 nick translation 106
氢键 hydrogen bond 55
轻链,L链 light chain 137
轻链启动子 light-strand promotor,*LSP* 137
倾向容错 error-prone 423,425
趋向性 divergence 356
去稳定元件 destablizing element 182
去阻抑 derepressed 244
全符合状态 in register 447
全酶 holoenzyme 107,156
缺倍体 nulliploid 43
缺失 deletion 39,426,441
缺体 nullisomy 43
确保裂解循环 ensuring lytic cycle 331

R

RI颗粒 RI particle 209
RNA编辑 RNA editing 285
RNA末端腺苷酸转移酶 RNA terminal riboadenylate transferase 180
RNA剪接 RNA splicing 184
RNA酶P RNase P 178
RNA酶D RNase D 179
RNA酶III RNase III 179
R带 R banding 23
ρ因子 ρ factor 168
染色粒 chromomere 11
染色粒间区 interchromomere 11
染色体 chromosome 9
染色体分带 chromosome banding 22
染色体畸变 chromosome aberration 38,441
染色体交叉 chromosome chiasmata 35,298
染色体组型模式图 idiogram 26
染色体组型图 karyogram 26
染色线 chromonema 9
染色质 chromatin 8
染色质小体 chromotosome 14
染色中心 chromocenter 20
酿酒酵母 *S. cerevisiae* 288
热激基因 heat shock gene 268
热激效应元件 heat shock response element,HSE 267
人类基因组计划 human genome project,HGP 94
熔点或熔解温度 melting temperature, T_m 58
熔解 melting 58
熔解曲线 melting curve 58
乳多孔病毒 papova virus 359
乳糖操纵子 lactose operon,*lac* 231
乳糖操纵子阻抑蛋白 *lac* repressor 232
弱化子 attenuator 258
弱化作用 attenuation 258
朊病毒 prion 50
朊病毒蛋白 prion protein,PrP 50

S

3-乙氧-2-氧丁醛 kethoxal 210
Sextama框 Sextama box 161
*SIR*基因座 selint information regulators,*SIR* 371
Southern印迹 Southern blot 92
SR蛋白 SR protein 189
SV40的大T抗原 SV40 large tumor antigen 145
SV40病毒 simian virus 40 144
S期 S phase 28
σ因子 σ factor 156
三联体 triad 64
三螺旋DNA或三链结构 triplex 64
三体性 trisomy 43
色氨酸合成酶 tryptophan synthetase 233
散布基因家族 interspersed gene family 87
上游 upstream 155
上游可读框 upstream open reading frame,uORF 288
上游调控元件 upstream control element, UCE 164
牛肌细胞 myogenic cell 278
生长点 growing point 101

生物膜 biomembrane 4
十字形结构 cruciform 61
释放因子 releasing factor 218
噬菌体展示 phage display 473
嗜热嗜酸菌 *Sulifolobus* 2
嗜盐菌 *Halobacterium* 2
收缩环 contractile ring 31
双骨髓瘤 double myelomas 397
双金属离子机制 two-metal-ion mechanism 112
双精受精 dispermy 43
双链断裂 double-strand break, DSB 406
双螺旋模型 double helix model 54
双三体性 ditrisomy 43
双微染色体 double-minute chromosomes 379
双线期 diplonema 35
顺反互补测验 *cis-trans* test 73
顺反子 cistron 74
顺接重复 tandem duplication 40
顺式剪接 *cis*-splicing 284
顺式显性 *cis*- dominance 235
顺式作用 *cis*-acting 229,325
顺式作用元件 *cis*-acting element 12,229
四聚体 tetramer 459
四链结构 tetraplex strucure 61
四体 tetrasomy 43
鼠伤寒沙门氏菌 *salmonella typhimurium* 259
松弛 DNA relaxed DNA 67
松弛型突变体 relaxed mutants, *rel* 257
随机扩增多态 DNA randomly amplified polymorphic DNA, RAPD 438
随体 satellite, trabant 11

T

TATA 框 TATA box 164
TATA 框结合蛋白 TATA binding protein, TBP 171
ter-Tus 复合物 *ter*-Tus complex 130
Tus 蛋白 terminus utilization substance 130
T 细胞抗原受体 T-cell receptors, TCR 386
T 分带法 terminal banding 24
τ 二聚体 τ dimmer 108
胎儿珠蛋白基因 foetal globin gene 35
唐氏综合征 Down syndrome 441
糖苷键 glucoside bond 53
糖皮质激素效应元件 glucocorticoid response element, GRE 267
特异性交换 specificity swap 272
特征序列 signature sequence 203
条件性突变 conditional mutation 429
调节基因 regulator gene 230
跳跃基因 jumping gene 322
铁离子调节因子 iron regulatory factor, IRF 289
铁离子应答元件 iron-responsive element, IRE 289
铁离子应答元件结合蛋白 iron-responsive element binding protein, IRE, BP 289
通用启动子 generic promoter 171
同翅目 Homoptera 10
同工 tRNA iso-accepting tRNA 202
同工型蛋白质 isoform 283
同义密码子 synonyms 200
同义突变 synonymous mutation 428
同源多倍体 autopolyploidy 43
同源基因 isogenes 376
同源框 homeobox 275
同源染色体 homologous chromosome 21
同源性重组 homologous recombination 290
同源性着色区 homogeneously staining region, HSR 432
同源域 homeodomain 270
同聚序列 homopolymeric stretch 110
头段 headpiece 239
凸环 bulge loop 60
突变 mutation 49,72
突变热点 mutation hotspot 74
突变生成作用 mutagenesis 427
突变体 mutant 72,427
脱落蛋白 split protein 209
拓扑异构酶 topoisomerase 69

W

5-氨基乙酰丙酸合成酶 5-aminolevulinate synthase, ALAS 289
5-溴脱氧尿嘧啶核苷 5-bromodeoxyuridine, BUdR 24
外被蛋白 coat protein 252
外切核酸酶 exonuclease 106
外显子 exon 76,183
外显子捕捉 exon trapping 93
外显子重排 exon shuffling 452
外周蛋白 extrinsic protein, peripheral protein 5
挽回系统 retrieval system 421
晚重组节 late recombination nodule 34
微球菌核酸酶 micrococcal nuclease 310
微卫星 DNA microsatellite DNA 89

卫星 DNA satellite DNA 89
位点特异性重组 site-specific recombination 291
温度敏感型突变体 temperature-sensitive mutant 429
无嘌呤或无嘧啶位点 apurnic/apyrimidinic sites, AP 409
无效等位基因 null alleles 430
无义密码子 nonsense codon 77
物理图谱 physical map 94

X

细胞核 nucleus 8
细胞膜 cell membrane 4
细胞器 organelle 5,90
细胞色素 c cytochrome c 461
细胞因子 cytokine 393
细胞质 cytoplasm 5
细胞质分裂 cytokinesis 30
细胞周期 cell cycle 27
细胞周期蛋白 cyclin 146
细线期 leptonema 33
θ-结构 θ-structure 101
匣式模型 cassette model 367
下游 downstream 155
下游启动子元件 downstream promoter element, DPE 166
线粒体 mitochondrion 90
线粒体终止因子 mitochondrial releasing factor, mRF 224
显性负调控 dominant negative 236
限制位点 restriction site 28
限制性片段长度多态性 restriction fragment length polymorphism, RFLP 438
诊断性限制性酶切位点 diagnostic restriction site 299
腺苷四磷酸 ppGpp 257
腺苷酸环化酶 adenylate cyclase 246
腺苷五磷酸 pppGpp 257
腺嘌呤 adenine 52
相对长度 relative length 25
交互易位 reciprocal translocation 41
镶嵌型 mosaicism 44
小沟 narrow groove 55
小结 nodes 298
小外显子 mini-exon 284
小卫星 DNA minisatellite DNA 89
效应元件 response element 267
协同合成 coordinating synthesis 125
协同调节 coordinate regulation 233
新着丝粒 neocentromere 10
信号传导 transduction 433
信号识别蛋白 signal recognition protein, SRP 402
信号识别颗粒 signal recognition particle, SRP 5,227
信号肽酶 signal peptidase 227
信使 RNA messenger RNA, mRNA 74
信息素 pheromone 364
性功能低下 hypogonadal, hpg 383
性染色体 sex chromosome 21
胸腺嘧啶 thymine 52
雄激素 androgen 273
秀丽隐杆线虫 *C. elegans* 285
血清效应元件 serum response element, SRE 267
序列均一性 sequence homogenization 451
瞬时表达 transient expression 167
瞬时转染 transient transfection 381

Y

1400 区 1400 region 211
亚倍体 hypoploid 43
亚病毒病原体 subviral pathogens 49
亚基因组 subgenomic 347
亚基装配缺失 subunit assembly defective 210
延迟运动 delayed movement, lagging 44
延伸因子 elongation factor 216
严重免疫联合缺陷病 severe combined immune deficiency, SCID 395
野生型 wild type 72,427
叶绿体 chloroplast 90
依赖 ρ 因子的终止子 rho-dependent terminator 168
移框突变 frameshift mutation 429
移码抑制 frameshift suppression 206
移位 translocation 217
移位易位 shift translocation 41
遗传密码 genetic code 48,74,199
遗传漂变 genetic drift 454
异丙基硫代-β-D-糖苷 isopropylthio-β-D-galactoside, IPTG 233
异常重组 illegitimate recombination 290,314
异染色质 heterochromatin 21
异位带 isochore 26
异源多倍体 allopolyploid 43
异周性 allocycle 21
抑制突变 suppressor mutation 429
抑制型 tRNA suppressor tRNA 204

义务诱导物 gratutous inducer 233
易位 translocation 39,441
印记 imprinting 282
应急调控 stringent control 256
应急因子 stringent factor 257
应急应答 stringent response 256
引发酶 primase 120
引发体 primosome 120
引发体组装位点 primosome assembly site, *pas* 124
引物 primer 110
荧光分带法 Q-banding 22
有丝分裂 mitosis 12,98
有丝分裂期 M phase 27
诱变 induced mutagenesis 427
诱导 induction 232
诱导物 inducer 233
诱导型 inducible 244
诱发突变体 induced mutant 427
预引发体 preprimosome 117,119
元件 element 266
原癌基因 proto-oncogene 432
起点 origin 100
起点识别复合物 origin recognition complex, ORC 148
原端粒 protelomere 11
原始细胞 primitive cell 2
原始液质 primeval soup 2
原噬菌体 prophage 310
跃进式复制 saltatory replication 445
运动失调性毛细血管扩张症 atacsia telangiectasia, AT 425

Z

杂交 hybridization 59
杂种不育 hybrid dysgenesis 342
再组装 reassociate 127
早中期 prometaphase 30
早重组节 early recombination nodule 34
增变基因 mutator gene 432,434
增强子 enhancer 266
增色效应 hyperchromic effect 58
增殖性细胞核抗原 proliferating cell nuclear antigen, PCNA 146
胀泡 puff 19
真端粒 eutelomere 11
真细菌 bacteriotida 3
整合 integrate 311
整合表达 integrant expression 167
整合蛋白 integral protein 5
整合酶 integrase, Int 311,347
整合宿主因子 integration host factor, IHF 311
整合体 intasome 314
正超螺旋 positive supercoil 67
正链 DNA plus strand DNA 348
正链病毒 plus strand virus 347
正向重复 direct repeats 324
正异固缩现象 positive heteropycnosis 21
支架附着区域 scaffold attachment region, SAR 16
支架环模型 scaffold-loop model 16
支链迁移 branch migration 295
脂质双分子层 lipid bilayer 4
脂质体 liposome 4
指令假说 instructive hypothesis 386
质膜 plasma membrane, plasmalemma 4
置换环或D环 displacement loop 124,137
中部核心结构 central core structure 183
中度重复 DNA moderately repetitive DNA 86
中间缺失 interstitial deletion 39
中期 metaphase 30
中心粒 centriole 7
中心体 centrosome 7
中央成分 central element 34
中央区 central space 34
中着丝粒 metacentric 10
终变期 diakinesis 35
终点 terminus 100
终止子 terminator 168,230
种 species 464
肿瘤生成 oncogenesis, Onc 352
肿瘤抑制基因 tumor suppressor gene, TS 433
轴心体 axial element 299
逐出区域 exit domain 210
主缢痕 primary constriction 9
转氨甲酰酶 transcarbamylase 380
转导病毒 transducing virus 352
转化 transformation 47
转化病毒 transforming virus 352
转化基因 transforming gene 353
转化基因组 transforming genome 353
转化因子 transforming factor 47
转换 transitions 428
转基因 transgenic 435
转录 transcription 155

转录单位 transcription unit 75,155
转录因子 transcription factor 230
转染 transfection 378
转铁蛋白 ferritin 289
转酯反应 transesterification 186
转座 transposition 88,290
转座酶 transposase 325
转座元件 transposable element 267,322,465
转位因子 transposable element 465
转座子 transposon 322
装配因子 assembly factor 174
着色性干皮病 xeroderma pigmentosum 425
着丝粒 centromere 9
着丝粒元件 centromere element, CEN 12
着丝粒指数 centromere index 25
子囊菌 *Ascomycetes* 300
自发突变生成 spontaneous mutagenesis 427
自发突变体 spontaneous mutant 427
自剪接 self-splicing, autosplicing 191
自杀底物 suicide substrate 312
自私DNA selfish DNA 323
自主复制序列 autonomously replicating sequence, *ARS* 13,147,370
自主元件 autonomous element 339
足迹法 footprinting 162
组成型表达 constitutive expression 268
组成型多倍体 constitutional polyploid 43
组成型基因表达 constitutive gene expression 234
组成型启动子元件 constitutive promoter element 268
组成型异染色质 constitutive heterochromatin 21
阻抑蛋白 repressor 230,450
阻抑型 repressible 244
阻抑作用 repression 233

A1C01860835B